

**THE INTEGRATION OF SAFETY AND HEALTH  
ASPECTS IN CHEMICAL PRODUCT DESIGN  
FRAMEWORK**

**JOON YOON TEN, MEng.**

**Thesis submitted to The University of Nottingham  
for the degree of Doctor of Philosophy**

**AUGUST 2017**

## **ABSTRACT**

Computer aided molecular design (CAMD) is a powerful technique to design molecules or chemical mixtures that fulfil a set of desirable target properties as specified by users. Molecular physical and thermodynamic properties are selected as the target properties to ensure that the designed molecules can achieve the property functionalities. However, the aspects of safety and health are not strongly emphasised as design objectives in many CAMD problems. In order to ensure that the synthesised molecule does not cause much harm and health-related risks to the consumers, it is critical to integrate both safety and health aspects as design factors in the current CAMD approaches.

The main focus of this research is to develop a novel chemical product design methodology that integrates the concept of inherent safety and occupational health aspects in a CAMD framework. The generated molecules that are optimised with respect to the target properties must be evaluated in terms of their safety and health performance. The assessment is conducted by safety and health-related parameters/sub-indexes that have significant adverse impact on both aspects. This proposed approach ensures that a product that possesses the desirable properties, and at the same time meets the safety and health criteria, is produced.

The next focus of this research is to generate optimal molecules with the desired functionalities and favourable safety and health attributes in a single-stage CAMD framework. Besides target properties, the concept of inherent safety and health is also considered as design objective to ensure that the synthesised molecules are simultaneously optimised with

regards to both criteria. Fuzzy optimisation approach is applied to optimise these two principal design criteria in this work.

As molecular properties are utilised as the parameters to examine the safety and health features of the molecules, these properties are often estimated through property prediction models. This research also focuses on the management of uncertainty resulted from properties used in the sub-indexes. The quantification of uncertainty helps to revise the safety and health measurement so that it can better reflect the inherent hazard level of the molecules.

The fourth focus of this research is to address the limitations present in the current method of molecular hazard quantification. The enhancement is carried out by adopting the ordered weighted averaging (OWA) operator method with the analytic hierarchy process (AHP) approach in the safety and health assessment. Two case studies on solvent design are considered to demonstrate the presented methodologies.

# ACKNOWLEDGEMENTS

I would first like to express my sincere gratitude and thanks to my principal supervisor, Dr. Nishanth Chemmangattuvalappil for his excellent support, guidance and patience throughout my PhD study and research in computer-aided molecular design (CAMD). I would also like to thank my co-supervisors, Prof. Denny Ng and Dr. Mimi Hassim, the latter of whom is from Universiti Teknologi Malaysia, for guiding my research and assisting me to develop my background in optimisation programming and inherent safety and occupational health. I am grateful to Prof. Raymond Tan and Dr. Michael Angelo Promentilla, both of whom are from De La Salle University, Manila, for introducing me with some mathematical and programming techniques that I adopted in my research. I appreciate the help from my former colleague, Dr Lik Yin Ng for his assistance in property prediction uncertainty. I would like to thank my fellow colleagues in my research group for creating the enjoyable working atmosphere and stimulating discussions during those years. I wish to acknowledge the research and technical support provided by the staffs of the University of Nottingham Malaysia Campus for aiding my research experience. The financial support (LRGS Grant, Project Code: LRGS/2013/UKM-UNMC/PT/05) provided by the Ministry of Science, Technology and Innovation (MOSTI) Malaysia is gratefully acknowledged.

Last but not least, I especially thank my parents and siblings for their continuous support and motivation throughout my PhD study and my life in general. Without their continued encouragement, this research work would never have come to fruition. I would also like to thank my friends for providing strength and friendship that I needed, and for motivating me to strive towards my goal.

# PUBLICATIONS

## Referred Journals

**Ten, J.Y.**, Hassim, M.H., Chemmangattuvalappil, N., Ng, D.K.S. (2016) A novel chemical product design framework with the integration of safety and health aspects. *Journal of Loss Prevention in the Process Industries*, 40, 67-80, DOI: 10.1016/j.jlp.2015.11.027

**Ten, J.Y.**, Hassim, M.H., Ng, D.K.S., Chemmangattuvalappil, N.G. (2017) A molecular design methodology by the simultaneous optimisation of performance, safety and health aspects. *Chemical Engineering Science*, 159, 140-153, DOI: 10.1016/j.ces.2016.03.026

**Ten, J.Y.**, Ng, L.Y., Hassim, M.H., Ng, D.K.S., Chemmangattuvalappil, N. (2017) Managing uncertainty on the integration of safety and health indexes in computer-aided molecular design. *Industrial & Engineering Chemistry Research*, 56, 10413-10427, DOI: 10.1021/acs.iecr.7b00768

**Ten, J.Y.**, Hassim, M.H., Promentilla, M.A.B., Tan, R.R., Ng, D.K.S., Chemmangattuvalappil, N.G. (2018) Enhancing molecular safety and health measurement via index smoothing and prioritisation. *Molecular Systems Design & Engineering*, DOI: 10.1039/C7ME00073A

## **Book Chapter**

**Ten, J.Y.**, Hassim, M.H., Ng, D.K.S., Chemmangattuvalappil, N.G. (2017) Chapter 7: The incorporation of safety and health aspects as design criteria in a novel chemical product design framework. *Computer Aided Chemical Engineering*, 39, 197-220, DOI: 10.1016/B978-0-444-63683-6.00007-1

## **Conference Proceedings**

**Ten, J.Y.**, Hassim, M.H., Ng, D.K.S., Chemmangattuvalappil, N. (2015) A novel chemical product design framework with the integration of safety and health aspects. *Chemical Engineering Transactions*, 45, 385-390, DOI: 10.3303/CET1545065

**Ten, J.Y.**, Hassim, M.H., Chemmangattuvalappil, N., Ng, D.K.S. (2015) A novel chemical product design framework with the integration of safety and health aspects. IN: *PRES Conference 2015*, Kuching, Malaysia.

**Ten, J.Y.**, Ng, L.Y., Hassim, M.H., Ng, D.K.S., Chemmangattuvalappil, N.G. (2015) Uncertainty analysis on the integration of safety and health indexes in a chemical product design framework. IN: *5<sup>th</sup> International Conference on Advances in Energy Research 2015*, Indian Institute of Technology Bombay, Mumbai, India.

**Ten, J.Y.**, Ng, L.Y., Hassim, M.H., Ng, D.K.S., Chemmangattuvalappil, N.G. (2016) Uncertainty Analysis on the Integration of Safety and Health Indexes in a Chemical Product Design. IN: *7<sup>th</sup> International Symposium on Design, Operation and Control of Chemical Processes*, The University of Tokyo, Tokyo, Japan.

# TABLE OF CONTENTS

<b>ABSTRACT</b>	i
<b>ACKNOWLEDGEMENTS</b>	iii
<b>PUBLICATION</b>	iv
<b>LIST OF TABLES</b>	xiii
<b>LIST OF FIGURES</b>	xvii
<b>NOMENCLATURE</b>	xix
<b>CHAPTER 1      INTRODUCTION</b>	<b>1</b>
1.1      Inherent Safety and Occupational Health	3
1.2      Computer-Aided Molecular Design (CAMD)	5
1.3      Research Gaps	7
1.4      Scopes of Research	11
1.4.1      Evaluate the safety and health aspects of optimal molecules using inherent safety and health indexes	11
1.4.2      Integrate inherent safety and health aspects into chemical product design framework	12
1.4.3      Manage uncertainty on the application of property prediction in safety and health sub- indexes	13
1.4.4      Improve the measurement of safety and health by introducing weight factors for sub-index prioritisation and smoothing sub-index scores	14
1.5      Research Objectives	15
1.6      Thesis Outline	16
1.7      Summary	16

<b>CHAPTER 2</b>	<b>LITERATURE REVIEW</b>	19
2.1	Introduction	19
2.2	Inherent Safety	20
2.2.1	Prototype Index for Inherent Safety (PIIS)	20
2.2.2	Inherent Safety Index (ISI)	22
2.2.3	<i>i</i> -Safe	23
2.2.4	Other Inherent Safety Indexes	23
2.3	Inherent Occupational Health	25
2.3.1	Prototype Index for Inherent Safety (PIIS)	26
2.3.2	Inherent Occupational Health Index (IOHI)	27
2.3.3	Applications of Inherent Safety and Occupational Health Indexes	28
2.4	Computer-Aided Molecular Design (CAMD)	30
2.4.1	Group Contribution (GC) Methods	31
2.4.2	Property Prediction Uncertainty	36
2.4.3	CAMD Applications	40
2.4.4	Multi-Objective Optimisation	50
2.5	Analytic Hierarchy Process (AHP)	54
2.6	Summary	57
<b>CHAPTER 3</b>	<b>RESEARCH METHODOLOGY</b>	60
3.1	Early Research Stage	60
3.2	Research Scope 1	60
3.3	Research Scope 2	61
3.4	Research Scope 3	62
3.5	Research Scope 4	63
3.6	Methodology Overview	64
3.7	Summary	65



<b>CHAPTER 4</b>	<b>A NOVEL CHEMICAL PRODUCT DESIGN</b>	69
	<b>FRAMEWORK WITH THE INTEGRATION OF</b>	
	<b>SAFETY AND HEALTH ASPECTS</b>	
4.1	Introduction	69
4.2	Problem Statement	70
4.3	Methodology	71
4.3.1	Problem Formulation	71
4.3.2	Model Development	72
4.3.3	Molecular Design	72
4.3.4	Optimisation Model	73
4.3.5	Performance Analysis	77
4.4	Case Study: Solvent Design for Gas Sweetening Process	83
4.4.1	Problem Formulation	84
4.4.1.1	Design Goal	84
4.4.1.2	Target Properties	85
4.4.2	Model Development	86
4.4.2.1	Property Prediction Model	86
4.4.3	Molecular Design	87
4.4.3.1	Molecular Blocks	87
4.4.3.2	Structural Constraints	88
4.4.4	Optimisation Model	88
4.4.4.1	Property Operator Targets	88
4.4.4.2	Fuzzy Optimisation	89
4.4.5	Performance Analysis	90
4.4.5.1	Selection of Inherent Safety and Health	90
	Sub-indexes	
4.4.5.2	Property Prediction Models	91
4.4.6	Results and Discussions	92

4.5	Summary	98
<b>CHAPTER 5</b>	<b>A MOLECULAR DESIGN METHODOLOGY BY THE SIMULTANEOUS OPTIMISATION OF PERFORMANCE, SAFETY AND HEALTH ASPECTS</b>	100
5.1	Introduction	100
5.2	Problem Statement	101
5.3	Methodology	102
5.3.1	Problem Formulation	102
5.3.2	Inherent Safety and Health Sub-index Selection	103
5.3.3	Model Development	104
5.3.3.1	Property Prediction Models	104
5.3.3.2	Disjunctive Programming for Allocation of Sub-index Scores	104
5.3.4	Molecular Design	108
5.3.4.1	Structural Constraints: First-Order Groups	109
5.3.4.2	Structural Constraints: Second-Order Groups	110
5.3.5	Optimisation Model	112
5.4	Case Study: Solvent Design for Gas Sweetening Process	113
5.4.1	Problem Formulation	113
5.4.2	Inherent Safety and Health Sub-index Selection	114
5.4.3	Model Development	114
5.4.4	Molecular Design	114
5.4.5	Optimisation Model	115

5.4.6	Results and Discussions	116
5.5	Summary	123
<b>CHAPTER 6</b>	<b>MANAGING UNCERTAINTY ON THE INTEGRATION OF SAFETY AND HEALTH INDEXES IN COMPUTER-AIDED MOLECULAR DESIGN</b>	125
6.1	Introduction	125
6.2	Problem Statement	126
6.3	Methodology	128
6.3.1	Managing Uncertainty in Sub-indexes	129
6.3.2	Allocation of Sub-index Scores with Disjunctive Programming	132
6.4	Case Study: Solvent Design for Extraction of Carotenoids	134
6.4.1	Problem Formulation	136
6.4.2	Selection of Inherent Safety and Health Sub- indexes	138
6.4.3	Model Development	139
6.4.3.1	Property Prediction Models	139
6.4.3.2	Managing Uncertainty in Sub-indexes	140
6.4.3.3	Allocation of Sub-index Scores with Disjunctive Programming	143
6.4.4	Molecular Design	144
6.4.5	Optimisation Model	144
6.4.6	Results and Discussions	147
6.5	Summary	151

<b>CHAPTER 7</b>	<b>ENHANCING MOLECULAR SAFETY AND HEALTH MEASUREMENT VIA INDEX SMOOTHING AND PRIORITISATION</b>	152
7.1	Introduction	152
7.2	Problem Statement	153
7.3	Analytic Hierarchy Process (AHP)	155
7.4	Ordered Weighted Averaging (OWA) Operator	159
7.5	Methodology	160
7.5.1	Problem Formulation and Safety and Health Assessment	161
7.5.2	Determination of Total Weighted Index Score	161
7.5.3	Smoothing Sub-index Scores	165
7.5.4	Allocation of Sub-index Scores with Disjunctive Programming	167
7.5.5	Molecular Design and Optimisation Model Formulation	169
7.6	Case Study: Solvent Design for Extraction of Carotenoids	169
7.6.1	Problem Formulation	169
7.6.2	Safety and Health Assessment	171
7.6.3	Smoothing Sub-index Scores	172
7.6.4	Molecular Design	181
7.6.5	Optimisation Model Formulation	181
7.6.6	Results and Discussions	182
7.7	Summary	189
<b>CHAPTER 8</b>	<b>CONCLUSIONS</b>	190
8.1	Achievements	190
8.2	Future Work	194

<b>REFERENCES</b>	198
<b>APPENDICES</b>	224
A.1    Lingo Coding (Chapter 4)	224
A.2    Lingo Coding (Chapter 5)	227
A.3    Lingo Coding (Chapter 6)	238
A.4    Lingo Coding (Chapter 7)	250

## LIST OF TABLES

<b>Table 2.1:</b>	Flammability scores (PIIS) (Edwards and Lawrence, 1993)	22
<b>Table 4.1:</b>	Parameters evaluated in inherent safety indexes	77
<b>Table 4.2:</b>	Flammability ( $I_{FL}$ ) sub-index (National Fire Protection Association, 2007)	79
<b>Table 4.3:</b>	Explosiveness ( $I_{EX}$ ) sub-index (ISI) (Heikkilä, 1999)	79
<b>Table 4.4:</b>	Parameters evaluated in the inherent health indexes	80
<b>Table 4.5:</b>	Viscosity ( $I_{\eta}$ ) sub-index (PRHI) (Hassim and Edwards, 2006)	81
<b>Table 4.6:</b>	Material phase ( $I_{MS}$ ) sub-index (IOHI) (Hassim and Hurme, 2010)	82
<b>Table 4.7:</b>	Volatility ( $I_V$ ) sub-index (IOHI) (Hassim and Hurme, 2010)	82
<b>Table 4.8:</b>	Exposure limit ( $I_{EL}$ ) sub-index (IOHI) (Hassim and Hurme, 2010)	82
<b>Table 4.9:</b>	Acute health hazard ( $I_{AH}$ ) sub-index (National Fire Protection Association, 2007)	83
<b>Table 4.10:</b>	Property targets for case study (gas sweetening)	86
<b>Table 4.11:</b>	GC models for selected properties in the case study	87
<b>Table 4.12:</b>	Property operator targets for case study (gas sweetening)	88
<b>Table 4.13:</b>	Property operator targets for design objectives (gas sweetening)	89

<b>Table 4.14:</b>	GC models for selected properties used in sub-indexes	91
<b>Table 4.15:</b>	Properties of MEA	93
<b>Table 4.16:</b>	The sub-index scores of MEA	93
<b>Table 4.17:</b>	The eight generated solvents with their properties (results for scope 1)	94
<b>Table 4.18:</b>	The eight generated solvents with their properties (continued) [results for scope 1]	95
<b>Table 4.19:</b>	The eight generated solvents with their sub-index scores (results for scope 1)	95
<b>Table 5.1:</b>	Property operator targets for design objectives (gas sweetening)	115
<b>Table 5.2:</b>	The generated solvents with their properties in Scenario A (results for scope 2)	119
<b>Table 5.3:</b>	The generated solvents with their properties in Scenario A (continued) [results for scope 2]	119
<b>Table 5.4:</b>	The generated solvents with their sub-index scores in Scenario A (results for scope 2)	120
<b>Table 5.5:</b>	The generated solvents with their properties in Scenario B (results for scope 2)	120
<b>Table 5.6:</b>	The generated solvents with their properties in Scenario B (continued) [results for scope 2]	121
<b>Table 5.7:</b>	The generated solvents with their sub-index scores in Scenario B (results for scope 2)	121
<b>Table 6.1:</b>	Property constraints for case study (carotenoid extraction)	137
<b>Table 6.2:</b>	GC models for selected properties in the case study (carotenoid extraction)	139

<b>Table 6.3:</b>	Viscosity ( $I_{\eta}$ ) sub-index (revised form from PRHI) (Hassim and Edwards, 2006)	141
<b>Table 6.4:</b>	Volatility ( $I_V$ ) sub-index (revised form from IOHI) (Hassim and Hurme, 2010)	142
<b>Table 6.5:</b>	Exposure limit ( $I_{EL}$ ) sub-index (revised form from IOHI) (Hassim and Hurme, 2010)	142
<b>Table 6.6:</b>	Acute health hazard ( $I_{AH}$ ) sub-index (revised form from National Fire Protection Association, 2007)	143
<b>Table 6.7:</b>	The six generated solvents with their properties (results for scope 3)	148
<b>Table 6.8:</b>	The six generated solvents with their sub-index scores (results for scope 3)	148
<b>Table 7.1:</b>	The fundamental AHP scale (Saaty, 1977)	156
<b>Table 7.2:</b>	Random consistency (RC) index (Saaty, 1980)	158
<b>Table 7.3:</b>	Pairwise comparison matrix	163
<b>Table 7.4:</b>	Smoothened $I_{FL}$ scores	177
<b>Table 7.5:</b>	Smoothened $I_{EX}$ scores	179
<b>Table 7.6:</b>	Smoothened $I_{\eta}$ scores	179
<b>Table 7.7:</b>	Smoothened $I_V$ scores	179
<b>Table 7.8:</b>	Smoothened $I_{EL}$ scores	180
<b>Table 7.9:</b>	Smoothened $I_{AH}$ scores	180
<b>Table 7.10:</b>	The six generated solvents with their properties (results for scope 4)	183
<b>Table 7.11:</b>	The six generated solvents with their sub-index scores (results for scope 4)	183
<b>Table 7.12:</b>	The six generated solvents (without index smoothing and prioritisation) with their sub-index scores [results for scope 4]	185



<b>Table 7.13:</b>	The six generated solvents with their properties (using alternate pairwise comparison matrix) [results for scope 4]	188
--------------------	---	-----

# LIST OF FIGURES

<b>Figure 3.1:</b>	Flow diagram of overall methodology in Research Scope 1	66
<b>Figure 3.2:</b>	Flow diagram of overall methodology in Research Scope 2	67
<b>Figure 3.3:</b>	Flow diagram of overall methodology in Research Scopes 3 and 4	68
<b>Figure 4.1:</b>	The degree of satisfaction, $\lambda_p$ curve for design objective to be maximised (a) or minimised (b)	74
<b>Figure 4.2:</b>	Simplified flow sheet of amine gas sweetening plant	84
<b>Figure 4.3:</b>	The best eight solvents with their molecular structures (results for scope 1)	94
<b>Figure 5.1:</b>	The graphical illustration of viscosity sub-index ( $I_\eta$ )	105
<b>Figure 5.2:</b>	The four possible scenarios where CH group is connected to (a) none, (b) one, (c) two or (d) three CH <sub>3</sub> group(s)	111
<b>Figure 5.3:</b>	The generated solvents with their molecular structures (results for scope 2)	118
<b>Figure 6.1:</b>	The allocation of sub-index scores based on property sub-ranges	126
<b>Figure 6.2:</b>	(a) Initial form of sub-index; (b) Revised sub-index with the incorporation of uncertainty	130

<b>Figure 6.3:</b>	(a) Initial form of sub-index; (b) Sub-index with the incorporation of uncertainty; (c) Revised sub-index with composite curve	131
<b>Figure 6.4:</b>	(a) Initial form of $I_{AH}$ sub-index; (b) $I_{AH}$ sub-index with the integration of uncertainty; (c) Revised $I_{AH}$ sub-index curve	141
<b>Figure 6.5:</b>	The generated solvents with their molecular structures (results for scope 3)	147
<b>Figure 7.1:</b>	Hierarchy for the weight determination of sub-indexes	162
<b>Figure 7.2:</b>	A scenario of two property values with different sub-index scores	165
<b>Figure 7.3:</b>	Revised sub-index with smoothened scores	166
<b>Figure 7.4:</b>	Sub-index scores of $I_{FL}$ when (a) $T_b < 37.8^\circ\text{C}$ ; (b) $T_b \geq 37.8^\circ\text{C}$	173
<b>Figure 7.5:</b>	Determining the range to be smoothened in $I_{FL}$ sub-index	174
<b>Figure 7.6:</b>	Smoothened $I_{FL}$ scores	175
<b>Figure 7.7:</b>	Combined $I_{FL}$ scores	177
<b>Figure 7.8:</b>	The generated solvents with their molecular structures (results for scope 4)	182

# NOMENCLATURE

## *Abbreviations*

2-MeTHF	2-methyltetrahydrofuran
AA	Arithmetic averaging
AAE	Average absolute error
AHP	Analytic hierarchy process
ANP	Analytic network process
ARE	Average relative error
BB	Branch-and-bound algorithm
BBIS	Bioenergy-based industrial symbiosis
CAISEN	Combined Approach for Inherent Safety and Environmental
CAMD	Computer-aided molecular design
CAOD	Computer-aided organic synthesis
CAPD	Computer-aided product design
CEI	Chemical Exposure Index
CFC	Chlorofluorocarbon
CHP	Combined heat and power
CI	Connectivity indices
CNN	Computational neural network
CoMT-CAMD	Continuous-molecular targeting computer-aided molecular design
COS	Cost of safety
COSMO-RS	Conductor-like Screening Model for Realistic Solvation
CPME	Cyclopentyl methyl ether
DEA	Diethanolamine
DGA	Diglycolamine

DIPA	Diisopropanolamine
DMC	Dimethyl carbonate
EACO	Efficient ant colony optimisation
EFB	Empty fruit bunch
EOS	Equations of state
FFB	Fresh fruit bunch
FMEA	Failure modes and effect analysis
FMFO	Fuzzy multi-footprint optimisation
GA	Geometric averaging
GC	Group contribution
GC <sup>+</sup>	Group contribution <sup>+</sup>
GHS	United Nations' Globally Harmonised System of Classification and Labelling of Chemicals
GRAND	Graphical Descriptive Technique for Inherent Safety Assessment
HAZOP	Hazard and operability analysis
HC	Health Code
HE	Health Effects
HHI	Health Hazard Index
HIRA	Hazard Identification and Ranking System
HMF	Hydroxymethylfurfural
HSP	Hansen Solubility Parameters
I2SI	Integrated Inherent Safety Index
IBI	Inherent Benign-ness Indicator
ICPHI	Inherent Chemical and Process Hazard Index
ICPRI	Inherent Chemical Process Route Index
ILO	International Labour Organization
IOHI	Inherent Occupational Health Index

IPA	Isopropyl alcohol
IS	Industrial symbiosis
ISD	Inherent safety design
ISI	Inherent Safety Index
IUPAC	International Union of Pure and Applied Chemistry
LB	Lower boundary value
LCA	Life cycle assessment
LOP	Layers of protection
MDEA	Methyldiethanolamine
MEA	Monoethanolamine
MHI	Material Harm Index
MIC	Methyl isocyanate
MILP	Mixed-integer linear programming
MINLP	Mixed-integer nonlinear programming
MSDS	Material safety data sheet
NFPA	National Fire Protection Association
NN	Neural network
NuDIST	Numerical Descriptive Inherent Safety Technique
ODP	Ozone depletion potential
OEL	Occupation Exposure Limit
OHHI	Occupational Health Hazard Index
OhHI	Occupational health Hazard Index
ORC	Organic Rankine cycle
OSHA	Occupational Safety and Health Administration
OWA	Ordered weighted averaging
PIIS	Prototype Index for Inherent Safety
PPF	Palm pressed fibre
PRHI	Process Route Healthiness Index

PRI	Process route index
QRA	Quantitative risk analysis
QSPR	Quantitative structure-property relationship
RMSE	Root-mean-square error
SD	Signature descriptors
SDE	Standard deviation
SHE	Safety, health and environment
SRLP	Successive regression and linear programming
TEA	Triethanolamine
TEST	Toxicity Estimation Software Tool
TI	Topological indices
UB	Upper boundary value
vdW	van der Waals
WAA	Weighted arithmetic averaging
WEC	Worker Exposure Concentration
WHO	World Health Organization
ZOLP	Zero-one linear programming

### *Parameters*

$A$	Pairwise comparison matrix
$a_{ij}$	Numerical scale to describe the relative importance of element $i$ to element $j$ in AHP
$A_{LD50}$	Universal constant for the GC model of $LD_{50}$
$B_{LD50}$	Universal constant for the GC model of $LD_{50}$
$C_{(CH_3)_2CH, id1}$	Boundary value to denote whether $(CH_3)_2CH$ is present in the molecule
$C_i$	Contribution of first-order group of type- $i$
$CI$	Consistency index

$CR$	Consistency ratio
$D_j$	Contribution of second-order group of type- $j$
$E_k$	Contribution of the third-order group of type- $k$
$F_{p0}$	Universal constant for the GC model of $F_p$
$g$	Coefficient to represent the type (acyclic, monocyclic, bicyclic or tricyclic) of compounds
$G_T$	Total number of groups needed to form the molecules
$H_{v0}$	Universal constant for the GC model of $H_v$
$I_{SHI}^L$	Lower bound of total index score
$I_{SHI}^U$	Upper bound of total index score
$K_{ow0}$	Universal constant for the GC model of $\log K_{ow}$
$L_{(CH_3)2CH}$	Lower limit for the number of $CH_3$ groups bonded to CH group
$LFL_{const}$	Universal constant for the GC model of $LFL$
$M$	An arbitrarily large number in modified max-min aggregation method
$n$	Number of elements in AHP
$p_L$	Lower bound of feasible $p$ value
$p_{switch}$	Property boundary value
$p_U$	Upper bound of feasible $p$ value
$R^2$	Coefficient of determination
$RI$	Random consistency index
$T$	Temperature
$T_{b0}$	Universal constant for the GC model of $T_b$
$T_{m0}$	Universal constant for the GC model of $T_m$
$U_{(CH_3)2CH}$	Upper limit for the number of $CH_3$ groups bonded to CH group
$UFL_{const}$	Universal constant for the GC model of $UFL$



$V_{gas,std}$	Molar volume of gas or vapour at standard condition (dm <sup>3</sup> /mol)
$v_i$	Valence of group $i$
$V_{m0}$	Universal constant for the GC model of $V_m$
$v_p^L$	Lower bound of target property
$v_p^U$	Upper bound of target property
$w$	Vector of priorities or weights
$w_i$	Weight of element of $i^{th}$ ordered position
$x$	Binary variable in general GC equation
$z$	Binary variable in general GC equation

### *Variables*

$b_{i_1,id_1,i_2,id_2}$	Binary variable denoting whether first-order group $i_1$ with id $id_1$ ( $i_1,id_1$ ) is connected to first-order group $i_2$ with id $id_2$ ( $i_2,id_2$ )
$b_{ij}$	Binary integer variable
$C_O$	Oxygen stoichiometric coefficient in a reaction
$F_p$	Flash point (°C)
$H_v$	Heat of vaporisation (kJ/mol)
$I$	Binary integer variable
$I_{(CH_3)_2CH,id_1}$	Binary variable which denotes whether group CH with $ID$ number $id_1$ is connected to at least two CH <sub>3</sub> groups
$I_\eta$	Penalty score for sub-index of viscosity
$I_{AH}$	Penalty score for sub-index of acute health hazard
$I_C$	Penalty score for sub-index of corrosiveness in IOHI
$I_{EL}$	Penalty score for sub-index of exposure limit
$I_{EX}$	Penalty score for sub-index of explosiveness
$I_{FL}$	Penalty score for sub-index of flammability

$I_{FL,low}$	Lower bound of $I_{FL}$ for a specific $F_p$
$I_{FL,up}$	Upper bound of $I_{FL}$ for a specific $F_p$
$I_{HH}$	Index for Health Hazards in IOHI
$I_{MS}$	Penalty score for sub-index of material phase
$I_P$	Penalty score for sub-index of pressure in IOHI
$I_{PM}$	Penalty score for sub-index of mode of process in IOHI
$I_{PPH}$	Index for Physical and Process Hazards in IOHI
$I_R$	Penalty score for sub-index of R-phrase in IOHI
$I_{SHI}$	Total index score of molecule
$I_{SHI,i}$	Total index score of molecule without index prioritisation and smoothing in Chapter 7
$I_{SHI,w}$	Total weighted index score of molecule
$I_T$	Penalty score for sub-index of temperature in IOHI
$I_V$	Penalty score for sub-index of volatility
$LC_{50}$	Acute toxicity (96-h) to fathead minnow (mg/l)
$LC_{50,inhalation}$	Acute inhalation toxicity
$LD_{50}$	Acute oral toxicity (rat) (mg/kg)
$LD_{50,dermal}$	Acute dermal toxicity
$LEL$	Lower explosion limit (vol%)
$LFL$	Lower flammability limit (vol%)
$\log K_{oc}$	Soil sorption coefficient
$\log K_{ow}$	Octanol-water partition coefficient
$M_{(CH_3)_2CH}$	Total number of $(CH_3)_2CH$ group present in the molecule
$M_j$	Frequency of second-order group of type- $j$
$M_w$	Molecular weight (g/mol)
$N_i$	Frequency of first-order group of type- $i$
$N_i^*$	Number of molecular group of type $i$ of the optimal solution(s)

$O_k$	Frequency of third-order group of type- $k$
$p$	Property value
$PEL$	Permissible exposure limit (ppm)
$R_a$	Distance of a solvent from the centre of the Hansen solubility sphere
$S$	Explosiveness ( $UEL - LEL$ ) (vol%)
$SI_{Ri}$	Sub-index with $i^{th}$ ordered position
$T_b$	Normal boiling point (°C)
$T_{b,diff}$	Difference in boiling point (°C)
$T_m$	Normal melting point (°C)
$UEL$	Upper explosion limit (vol%)
$UFL$	Upper flammability limit (vol%)
$V_m$	Liquid molar volume (cm <sup>3</sup> /mol)
$V_p$	Target property value
$VP$	Vapour pressure (mmHg)
$W_s$	Aqueous solubility
$Z_{i1,id1}$	Binary variable representing the existence of group ( $i_1, id_1$ ) in molecule

### *Greek Symbols*

$\delta$	Wideness of the smoothened region
$\delta_d$	Dispersion solubility parameter of HSP
$\delta_h$	Hydrogen bonding solubility parameter of HSP
$\delta_p$	Polar solubility parameter of HSP
$\eta$	Viscosity (cP)
$\lambda$	Degree of satisfaction for least satisfied objective
$\lambda_I$	Degree of satisfaction for total index score of molecule
$\lambda_{max}$	Maximum eigenvalue in AHP

$\lambda_p$	Degree of satisfaction for target property $p$
$\Omega_p$	Property operator for target property $p$
$\sigma_p$	Standard deviation for property $p$

# **CHAPTER 1**

## **INTRODUCTION**

When dealing with product and process plant design, several decisions have to be considered during the research and development stage. These include determining the composition of chemical mixtures to achieve the desired product properties, constructing plant flowsheets and identifying the suitable operating conditions of the process (Seider et al., 2004). In a conventional design problem, the engineering technical aspects and economic factors are often the two key components considered as the decision-making criteria. Over the past few decades, many chemical processing industries have widely adopted the concept of sustainability development. Different aspects of sustainability such as environment, health and safety have received increasing attention in process plant development and design (Rathnayaka et al., 2014). Nowadays, process safety is regarded as a vital decision-making component especially in the chemical and petrochemical industries (Ee et al., 2015). This is due to the constant increasing scale of industrial operations and growing amount of industrial accidents reported in the process industries (Khan and Amyotte, 2004). Most accidents have occurred due to the mishandling of hazardous materials, combustible dusts and reactive chemicals (Chen et al., 2015). For instance, the Bhopal disaster, which was considered as one of the worst chemical disasters in history, has resulted in at least 3,800 fatalities immediately after the incident. The cause was due to the leakage of more than 40 tons of methyl isocyanate (MIC) gas, which is extremely toxic to living beings, from the pesticide plant (Broughton, 2005). Besides, the 1974 catastrophic explosion in a chemical plant close to the village of

Flixborough also served as an initiator for thorough reviews on methods to enhance the safety factors in the chemical process industry (Hansson, 2010).

Besides process safety, another major concern in process industries is the occupational health of the workers. As reported by the International Labour Organisation (ILO), over two million people around the world die from work-related diseases annually while occupational accidents have resulted in three hundred thousand fatalities yearly. Since more death tolls have been caused by work-related illness, the importance of occupational health in the industry must be treated as equally significant as process safety in process plants. One way to eliminate or minimise hazards in the plant is to operate the process with milder conditions and substitute those hazardous chemicals with less harmful ones. Any unintentional release or leakage of safer chemicals will not cause much adverse impacts to the people and environment. This concept of 'embedding' safety by eliminating or minimising process hazards is known as inherent safety, which was first introduced by Trevor Kletz (Kletz, 1978). Inherent safety design (ISD) serves differently compared to conventional safety concept as it strives to eliminate or minimise hazards through the implementation of inherent safety principles in plant design. The concept of inherent safety also assisted the development of a new health assessment concept, known as inherent occupational health (Hassim and Edwards, 2006). Both inherent safety and inherent occupational health share common key principles that include avoiding the use of hazardous materials and operating simpler processes with milder process conditions (Kletz, 1984; Hassim and Hurme, 2010a). The former principle serves as the main motivation of this research, thus there is a need to incorporate the inherent safety and

occupational health assessment while deciding on the use of less harmful chemicals in the process.

## **1.1 Inherent Safety and Occupational Health**

Conventionally, plant hazards are managed through the establishment of hazard identification and analysis techniques such as failure modes and effect analysis (FMEA), quantitative risk analysis (QRA) and hazard and operability analysis (HAZOP) (Palaniappan et al., 2002). These techniques assist in controlling hazard in order to minimise the consequences of a possible accident through the installation of add-on technological safety barriers, such as emergency shutoff valves, automatic interlocking devices, and flammable gas detector (Kang et al., 2016), which act as parts of layers of protection (LOP). However, the installation of add-on barriers often complicates plant design and further increases the capital costs (Srinivasan and Nhan, 2008). In fact, several past accidents such as the Flixborough (1974) and Piper Alpha (1987) catastrophes have occurred due to the failure of LOP in controlling hazards. This is because hazards remain present in the system since safety barriers do not eliminate them (Abidin et al., 2016). Rather than controlling hazards, the alternative to address these safety concerns is by eliminating them at their sources (Mansfield and Hawksley, 1998). After the Flixborough disaster in 1974, Trevor Kletz has presented a lecture titled "What You Don't Have, Can't Leak," which was deemed the first time that the concept of ISD in chemical processes has been formally introduced. Rather than applying the traditional safety concept to manage hazard, Kletz proposed that the process should be modified in a manner that hazard can be completely eliminated or its magnitude and likelihood of occurrence can be significantly minimised. ISD is best employed during the early stage of

design, particularly the research and development phase, when decision-making on process concept and reaction chemistry are made. This is due to the great deal of freedom in the selection of chemistry, solvents, raw materials, intermediates, unit operations, plant location and process parameters in early stage. Hence, it has the greatest potential opportunities for impacting the risk profile (Srinivasan and Natarajan, 2012). As the design progresses to later stages, the opportunities to modify the inherent safety of a process decrease as more engineering and financial decisions have already been decided in earlier stages (Heikkilä, 1999). Kletz (1991) has listed the Basic Principles of Inherent Safety, which comprise of intensification, substitution, attenuation, simplification, etc. The principle of intensification involves minimising the inventory of hazardous materials which lowers the magnitude of adverse impact to the workers due to leaks. The principle of substitution promotes the replacement of hazardous materials with less harmful ones. The principle of attenuation proposes the process operations to be conducted under milder and less hazardous conditions. Meanwhile, the principle of simplification aims to eliminate unnecessary complexity in process design to minimise the likelihoods for process error and wrong operation.

Besides process safety, the concern of occupational health has also received much attention even though it is not discussed as extensively as process safety. According to the World Health Organization (WHO), occupational health strives to improve working conditions and other aspects that are related to environment hygiene. Generally, health impacts differ from safety impacts in many aspects. The main distinction is, safety hazards may result in acute effects, whereas health impacts often deal with chronic diseases, due to prolonged exposure. Since the effects of health hazards can only be observed after long duration, this has resulted in



health-related events to receive much less attention as compared to safety. However, more fatalities are reported to be caused by work-related diseases than that of industrial accidents (Wenham, 2002). Hence, the improvement of occupational health in the industry has the same importance as the enhancement of process safety in process plants. This has prompted the establishment of inherent occupational health, in which the principal goal is to minimise the occupational health hazards arising from chemical processes to the workers (Hassim and Hurme, 2010a).

As mentioned previously, mishandling of hazardous materials is one of the key factors that result in industrial tragedy. One of the mutual aims and principles of both inherent safety and inherent occupational health is to promote the substitution of hazardous materials used in the plants with less dangerous ones. For instance, a highly toxic chemical involved in the process can be substituted with a less or non-toxic chemical. However, it is crucial that the latter chemical must demonstrate compatible product performance similar to that of the former. The potentially safer chemical can be a conventionally used chemical or a new chemical which has not been commercially available. An approach known as the computer-aided molecular design (CAMD) technique can be applied for the exploration and design of molecules that possess desirable functionalities and characteristics. The idea of integrating inherent safety and occupational health into CAMD will serve as the core objective of this research work.

## **1.2 Computer-Aided Molecular Design (CAMD)**

Traditionally, the exploration for new chemicals often couples with trial and error approach where numerous molecular compounds are synthesised in the laboratory. The properties of the generated compounds

are then examined based on the analytical results to validate whether the desired requirements are attained. However, this trial-and-error bottom-up method does not perform effectively in the present increasingly demanding global business environment as it is time-consuming and costly (Gebreslassie and Diwekar, 2015). On the other hand, another promising approach known as the chemical product design technique can greatly assist in the search of a potential chemical candidate with reduced time and effort (Duvedi and Achenie, 1996). In order to develop a chemical product that satisfies a specified design problem, Cussler and Moggridge (2001) have proposed four primary steps in their design process: define needs, generate ideas to fulfil needs, select among ideas, and manufacture final chemical product. According to Achenie et al. (2003), the second and third steps of the chemical product design process by Cussler and Moggridge (2001) can be combined to represent a molecular design problem. In order to solve the problem, numerous methods such as empirical trial and error approaches, mathematical programming, and hybrid techniques are often utilised. In the case where appropriate property models are available, the molecular design problem can be converted into a CAMD problem, which is a computer-aided based procedure employing property models. A CAMD problem is also equivalent to a 'reverse property prediction' problem, in which the product needs are given in terms of physicochemical properties, but the chemical identities (molecular or atomic structure) or their mixture (compositions) are unknown. In order to determine the chemical structures that best fulfil the needs, the reverse engineering approach coupling with CAMD technique (Heintz et al., 2014a) can be adequately employed. CAMD has been recognised as a powerful tool to identify molecules having a desirable set of physicochemical properties (Harper and Gani, 2000). Existing databases comprising of many chemical groups or molecular building blocks are

utilised to explore a large amount of conventional or novel molecular structures that can satisfy the product needs of interest. As the chemical identities or their mixture are unknown, property models may be employed in a generate and test solution approach by coupling with CAMD where the property prediction problem is solved iteratively to test the generated alternatives (Gani, 2004a). The problem can be formulated using a mathematical programming model to generate the optimal solution (Churi and Achenie, 1996). With this advantage, CAMD can be extensively used for the design of economically and sustainability superior chemical agents for industrial or other purposes. In this research, CAMD technique is employed to synthesise molecules that do not only achieve the specified target properties, but also display favourable safety and health characteristics. The concept of inherent safety and occupational health is integrated into the CAMD framework to serve as an assessment tool to ensure that molecules with high performance in terms of safety, health and product functionality are generated.

### **1.3 Research Gaps**

In CAMD problems, the development of a product is based on the needs of the customers. These needs are expressed in terms of product specifications, which are often represented by the physicochemical properties. The molecules generated must meet these target properties to ensure that the synthesised molecules can function and behave in the desired manner. Over the decades, the aspects of safety and health have been included as design constraints in molecular design problem. Molecules that do not meet the safety and health constraints are therefore screened out. However, the implementation of such constraints may suppress the generation of molecules which perform dominantly with respect to product

functionality, but are excluded due to not complying with the imposed constraints. There exists this major research gap in which a CAMD framework should also conduct a trade-off between the advantages in the performance of the excluded molecules and their safety attributes (Lampe et al., 2015). Instead of solving the CAMD problem with respect to only optimising the targeted functional properties, the safety and health aspects of the molecules must also receive the same emphasis. Hence, there exists a necessity to integrate safety and health aspects as design objectives along with target properties to ensure that the designed molecule is simultaneously optimised with respect to safety, health and functional performance. Over the years, chemical industries have played a crucial role in promoting sustainable development due to the possible impact on the environment, health and safety of its product and process activities (Heintz et al., 2014b). Mansfield and Hawksley (1998) highlighted the importance of integrating inherent safety, health and environment (SHE), rather than regarding each of these three factors in isolation as they are all interrelated.

In order to assess the level of inherent SHE of a process route, index-based metrics are often applied to serve as indicators for the evaluation (Edwards and Lawrence, 1993; Cave and Edwards, 1997; Heikkilä, 1999; Hassim and Hurme, 2010a). However, most of the established inherent indexes are developed to examine a single aspect; either inherent safety, inherent occupational health or inherent environmental hazard in a process route. Nevertheless, the methodology used to develop these indexes is similar, in which parameters that might significantly contribute to adverse safety, health or environmental impacts are first identified. However, not all identified parameters can be included in the indexes as the selection of parameters is often confined by the

availability of information during early design stage. The chosen parameters can be categorised into two classes, namely parameters that are linked to the chemicals (e.g. molecular physical and thermodynamic properties) and parameters that are associated to the process conditions. Each parameter is also known as a sub-index, which is assigned with a sub-index value or penalty score depending on the degree of potential hazards or the probability of exposure to the hazards. Most inherent safety and occupational health indexes are rather straightforward to be applied, as the inherent safety or health level of a route can be quantified by the summation of sub-index values of all contributing sub-indexes. The route with a lower sum is deemed to possess lower risk. On the other hand, inherent environmental indexes are more complicated as additional calculations are needed in order to determine the overall index value. Therefore, this research work only focuses on the application of inherent safety and occupational health indexes. Since most inherent safety and occupational health indexes are presented for process route design and selection, sub-indexes related to process conditions are not suitable to be incorporated in a molecular design problem. Hence, only chemical-related sub-indexes are considered, and they are generally assessed by physicochemical properties. Most CAMD methods have depended on the group contribution based property prediction methods to estimate the properties of molecules. These property prediction methods are rather user-friendly, and they have acceptable accuracy in estimating properties of simple-structured molecules (Harper et al., 1999). Thus, chemical-related sub-indexes can be integrated into a CAMD problem to evaluate the safety and health performance of the generated molecules.

The application of property prediction methods in calculating properties offers the advantage of swift property estimation without the

need of extensive computational resources. However, the main concern for the application of property prediction model is the accuracy of the estimated property values may not always be guaranteed as there are commonly 5-10% (or higher) of discrepancies between the actual experimental values and property predictions (Maranas, 1997). Since the sub-index values rely strongly on the property values, thus the accuracy of sub-index values assigned to the molecules to quantify their corresponding inherent hazard level is contributed by the accuracy and reliability of property prediction models. The deviation of predicted value may result in the shift of sub-index score to an inaccurate value, thereby incorrectly reflecting the inherent hazard level. Hence, there exists this research gap where uncertainty resulted from the application of property prediction methods must be managed in order to determine the inherent safety and health sub-index values with improved accuracy to better represent the intrinsic hazard level of the generated molecules under uncertainty condition.

In each sub-index, the hazard level is generally measured by one or more physicochemical properties of the molecules. For instance, the sub-index of volatility is being examined using the normal boiling point, while the sub-index of flammability is assessed by both flash point and boiling point. All sub-indexes share a common trend, in which the properties are divided into multiple sub-ranges, and a single penalty score is then assigned to each sub-range. A sub-range with higher risk level receives a higher penalty score and vice versa. Through this score allocation method, property values that fall within the same sub-range are considered to exhibit similar hazard level. At the property boundary which separates two adjacent sub-ranges, the penalty score switches abruptly from one value to another. The main drawback of this current allocation method is the

discontinuity at the boundary, which may distort the comparison of alternatives that are near these limits. In order to address this weakness, the scorings near the boundary should be smoothened to ensure continuity in allocating scorings at all property range. Besides, most inherent safety and occupational health indexes presented that the quantification of the overall risk level of a process route is conducted by summing up all the sub-index values involved in the evaluation. The proposed approach treats all sub-indexes to have equal impact to the plant safety and occupational health of the workers. However, some sub-indexes may bring about higher adverse impact as compared to others. This issue leads to another research gap where a systematic procedure must be developed to improve the current approach of determining the overall hazard level of molecules by introducing weight factor to each sub-index. This ensures that a higher impact sub-index is penalised with higher degree, so that a more conservative solution can be formulated.

## **1.4 Scopes of Research**

Based on the problems identified in Section 1.3, the scopes of this research work are defined as follow:

### **1.4.1 Evaluate the safety and health aspects of optimal molecules using inherent safety and health indexes**

CAMD technique is employed for designing molecules that meet the desirable functionality and attributes, which can be expressed in terms of targeted physicochemical properties for which the molecules must fulfil. In a CAMD problem, these properties are commonly estimated through group contribution based property prediction models, which can be integrated

into the mathematical optimisation model. A set of molecules that satisfies these target properties is generated by the optimisation model. These molecules then undergo performance analysis, mainly focusing on the aspects of safety and health to serve as the assessment criteria in order to identify the best solution. Some well-established inherent safety and health indexes are applied to serve as the evaluation tool, by utilising several significant safety and health sub-indexes to measure the molecular performance. The safety and health level of each molecule is determined and quantified by the total sub-index value allocated to it. The molecules are then compared and ranked by their corresponding total index scores, which represent the overall hazard level exhibited by the molecules. The molecule with the lowest risk level is then selected as the preferable solution.

#### **1.4.2 Integrate inherent safety and health aspects into chemical product design framework**

Based on the previous scope in Section 1.4.1, the target properties given by physical and thermodynamic properties are usually the key design criteria that must be attained in order to satisfy the needs detailed by customers. Due to increasing attentions on the concept of sustainability in recent decades, the aspects of safety and health are mostly considered as design constraints so that molecules that do not meet such criteria are screened out. However, the implementation of such constraints may suppress the generation of molecules which demonstrate superior performance in terms of product functionality, but are eliminated for not satisfying the constraints. In this scope, inherent safety and health of the molecule are served as the design objectives, along with the targeted properties of the molecule. The CAMD problem now becomes a multi-



objective design problem where molecule has to be optimised in terms of target property functionalities and safety and health performance. This provides greater flexibility as trade-off between functional performance of molecules and their safety and health attributes is considered. Fuzzy optimisation is employed to solve this multi-objective optimisation problem. To carry out the safety and health assessment, inherent safety and health sub-indexes are applied to identify sub-index values that will indicate the molecular hazard level. In order to simultaneously determine the optimal molecular structure and its corresponding sub-index values, disjunctive programming techniques are used to convert the molecular physicochemical properties into their respective scores.

#### **1.4.3 Manage uncertainty on the application of property prediction in safety and health sub-indexes**

The allocation of safety and health sub-index scores to the molecules is highly dependent on their physicochemical properties. Since the identity of the molecules is initially unknown, and the goal of a CAMD problem is to identify the optimal molecular structures, property prediction methods are commonly coupled in such problem to assist in estimating the properties. The accuracy of the scores assigned to the molecules relies heavily on the accuracy of the property values as estimated by the prediction models. However, it is common that most prediction models have 10% or higher discrepancies between the actual experimental values and predicted values. The discrepancies may result in the inaccuracy on the allocation of scores to the molecules. Thus, uncertainty resulted from the application of property prediction models to determine the sub-index scores must be managed. The goal is to improve the scorings of safety and

health sub-indexes to better reflect the inherent safety and health level of different molecules under property prediction uncertainty.

#### **1.4.4 Improve the measurement of safety and health by introducing weight factors for sub-index prioritisation and smoothing sub-index scores**

Based on the current inherent safety and health indexes taken from literature, each sub-index is measured using single or multiple molecular properties, and subjective scaling is employed where properties are first divided into a several sub-ranges. Each sub-range is given a single penalty score to represent a certain degree of hazard level. A property boundary is the point that separates two adjacent sub-ranges, and at this point the score switches abruptly from one value to another. Two molecules with property values falling within the same sub-range are assigned identical penalty score and are both considered to possess similar degree of hazard with respect to that particular sub-index. However, when another two molecules have property values that are close to one another but are separated by the property boundary, they each fall in an adjacent sub-range, thus resulting in the allocation of different penalty scores to each molecule. This scenario brings about the main weakness of this score allocation approach, which is the discontinuity in scorings at the boundary. In this scope, a modification is made on the scorings at the property boundary region by smoothing them to ensure that there is always continuity in terms of scorings at any feasible property value.

Besides, the overall hazard of a molecule is quantified by summing all the sub-index values that are involved in assessing the inherent safety and health level of the molecules. Through this method, all sub-indexes are

considered to have equivalent contribution to the overall adverse safety and health impacts. In this scope, weight factor is introduced to each sub-index to distinguish the level of importance among sub-indexes. A higher impact sub-index is given a higher weight to 'magnify' the hazardous condition of that particular sub-index. One method that can be utilised to determine the weight factors is the analytic hierarchy process (AHP), which is a structured multi-attribute decision approach. Since weaker-performing sub-indexes (higher penalty) are prioritised (given higher weight), this approach of determining the overall hazard ensures that the CAMD optimisation model is able to generate a more conservative solution with regards to safety and health performance.

## **1.5 Research Objectives**

The scopes identified in Section 1.4 lead to the objectives of this research, and they are shown as follow:

1. To apply inherent safety and health indexes for evaluating the safety and health attributes of optimal molecules generated by CAMD techniques.
2. To integrate safety and health aspects as design objectives in CAMD framework with the application of inherent sub-indexes.
3. To manage uncertainty resulted from property prediction on inherent safety and health sub-indexes.
4. To improve the measurement of molecular safety and health performance by introducing weight factors to the sub-indexes and smoothing the sub-index scores.

## **1.6 Thesis Outline**

The outline of this thesis is organised as follows. Chapter 2 presents the literature review on inherent safety, inherent occupational health, the application of inherent indexes, CAMD and its applications, property prediction methods employed in CAMD, multi-objective optimisation approach, and AHP to manage decision-making problem. Chapter 3 depicts the methodology of the four proposed research scopes. The four scopes are then demonstrated in Chapters 4 to 7. The first research scope is presented in Chapter 4, which is to perform a safety and health assessment on the generated optimal molecules with the application of safety and health sub-indexes. The second research scope is illustrated in Chapter 5, which aims to consider both safety and health aspects as design objectives to generate molecules that not only exhibit the desired target property functionalities, but also display favourable safety and health characteristics. The third research scope is demonstrated in Chapter 6, which helps to manage uncertainty resulted from the application of property prediction models in the sub-indexes. The fourth research scope is proposed in Chapter 7, which smoothens the sub-indexes at the property boundary region, and introduces weight factors to all sub-indexes to reflect the impact level of different sub-indexes.

## **1.7 Summary**

In Chapter 1, the importance of safety and health aspects in process plants has been highlighted. Most industrial accidents are resulted by the mishandling of hazardous and reactive chemical substances. One way to address such issue is to substitute those dangerous chemicals with less harmful ones. The search for new replacement chemicals can be conducted by the application of CAMD, which is an effective tool to explore suitable

chemical candidates that exhibit the desired target properties as specified by users. Through CAMD, it is possible to look for existing molecules or novel molecular structures that not only demonstrate compatible property functionalities similar to those of the substituted hazardous chemicals, but also with improved safety and health characteristics. In order to ensure that both safety and health aspects are taken into consideration in CAMD, the concept of inherent safety and occupational health can be adopted in the framework. This concept aims to eliminate or reduce safety and occupational-related hazards rather than controlling them. Many inherent safety and occupational health indexes have been established to quantify and compare hazards in different chemical process routes for the synthesis of a specific product. The indexes are comprised of several safety and health parameters/sub-indexes, which employ molecular properties to measure the degree of potential hazards with respect to different attributes.

As discussed earlier, the first research scope is to utilise chemical-related sub-indexes to assess the safety and health aspects of the molecules generated from CAMD model. This ensures that safer molecules with better safety and health characteristics are selected to substitute the usage of dangerous chemicals in process plant. The research gaps for the management of safety and health criteria in the current state of CAMD are highlighted. Most CAMD works implemented safety and health criteria as design constraints, hence molecules that do not meet such constraints are screened out. The second scope of this research is to conduct a trade-off between the property functionalities and the safety and health performance of the molecules. As there are several design objectives in the CAMD model, fuzzy optimisation formulation can be adequately applied to simultaneously optimise all objectives. This ensures that the solutions

generated are optimised with respect to functional performance, safety and health aspects.

The method to quantify molecular hazard is by adopting sub-indexes to assign penalty scores based on the safety and health-related properties. As CAMD requires the application of property prediction models to estimate molecular properties, this can result in the issue of uncertainty when utilising safety and health sub-indexes to evaluate the molecules. Therefore, the third research scope is to account for property prediction uncertainty in the sub-indexes, so that the sub-index scores can be enhanced to better reflect the hazard level exhibited by the molecules. As these sub-indexes are developed in such a way that properties are divided into multiple sub-ranges with different scorings to represent different hazard level, the fourth research scope is to improve the index scaling by smoothing the scores at the property boundaries. This modification is carried out to address the limitation of the existing scoring approach and ensure continuity in terms of scores allocation. Besides, this scope also covers the introduction of weight factors to the sub-indexes, so that sub-indexes with higher impact are prioritised to provide greater contribution to the overall hazard level of the molecules.

## **CHAPTER 2**

### **LITERATURE REVIEW**

#### **2.1 Introduction**

The core objective of this research work is to present a novel CAMD framework by implementing the aspects of safety and health as design objectives. This ensures that the CAMD model is able to synthesis molecules that fulfil the targeted properties and also exhibit acceptable safety and health characteristics. The application of such molecules in process plants will minimise the magnitude of harm to the people and surrounding community. The concepts of inherent safety and occupational health are integrated into the CAMD framework to carry out the molecular safety and health evaluation. Section 2.2 discusses the overview of inherent safety, while also reviews some of the prominent inherent safety indexes in literature. Meanwhile, Section 2.3 examines the background of some early health indexes, and also the development of inherent indexes specifically on occupational health. Section 2.4 then presents the introduction to CAMD, and the application of property prediction methods is also discussed. Section 2.4.1 provides a background of some of the commonly used property estimation methods in CAMD. Section 2.4.2 details the issues of property prediction uncertainty, and the existing works that have been carried out to manage uncertainty of the property data. The applications of CAMD in the design of numerous chemical products are illustrated in Section 2.4.3, while Section 2.4.4 illustrates the optimisation technique utilised to solve a multi-objective design problem. Section 2.5 provides brief explanation on analytic hierarchy process (AHP), which is a

tool to manage multi-attributes decision-making problems. The employment of AHP in many applications is also reviewed.

## **2.2 Inherent Safety**

A process is deemed to be inherently safer when the amount of hazardous materials and operations involved is minimised. An inherently safer plant is more appealing than a conventional plant as the former has reduced 'built-in' hazard potential (Rahman et al., 2005) and less add-on protective systems that leads to process simplification (Hassim and Hurme, 2010a). The concept of inherent safety has long been implemented in many applications, such as process concept evaluation, process route planning and plant layout design (Okoh and Haugen, 2014).

Generally, the best way to integrate inherent safety principles to their full extent is to employ them in the earlier design phase. In order to identify the level of inherent safety for different process routes, one commonly used method is the index-based approach, where each process route is evaluated by several safety factors or parameters. It is able to swiftly produce reliable results that can assist users in deciding process route with better safety attributes (Gnoni and Bragatto, 2013). In the past few decades, numerous inherent safety indexes have already been established, which began with the introduction of Prototype Index for Inherent Safety (PIIS) as developed by Edwards and Lawrence (1993). The following section provides a full background on the development of PIIS.

### **2.2.1 Prototype Index for Inherent Safety (PIIS)**

PIIS is a simple index-based method to quantify the inherent safety of a process route, which mainly focuses on reaction step. The



development of this index is motivated by the existing index scoring schemes of Dow Fire and Explosion Index and Mond Fire and Explosion Index. Edwards and Lawrence (1993) first proposed a preliminary list of safety-related parameters that could be integrated into the index to assist with the safety assessment, which comprises of inventory, phase, temperature, pressure, heat of reaction, new phase generation, catalysts, side reactions, waste products, reaction yield, reaction rate, viscosity, flammability, explosiveness, corrosiveness, and toxicity. Since PIIS aims to measure the inherent safety of a route during conceptual design stage, the selection of parameters is limited by the data availability at that stage. Due to this reason, seven parameters are shortlisted to form the index scoring scheme, which can be classified into chemical score and process score. Chemical score is related to properties of chemicals involved in the reaction step, which comprise of inventory, flammability, explosiveness, and toxicity. Meanwhile, process score is associated to the reaction conditions, which cover temperature, pressure, and yield. The parameter scores are formulated by dividing the domain of values for a parameter into intervals and allocating a score to each interval. For instance, the flammability scores provided by PIIS are given in Table 2.1. This parameter is assessed by two properties, namely flash point ( $F_p$ ) and boiling point ( $T_b$ ). As shown in Table 2.1, these properties are divided into few intervals, where each interval is given a discrete score to represent the degree of hazard. The scores are assigned in a manner that higher score represents higher safety hazard level. The total score for each reaction step is the summation of chemical and process scores, while the total index value of a process route is determined by summing up the scores received by each reaction step in the process route.

**Table 2.1:** Flammability scores (PIIS) (Edwards and Lawrence, 1993)

Flammability	Score
Non-combustible	0
$F_p > 140^\circ\text{F}$	1
$100^\circ\text{F} < F_p < 140^\circ\text{F}$	2
$F_p < 100^\circ\text{F} \ \& \ T_b > 100^\circ\text{F}$	3
$F_p < 100^\circ\text{F} \ \& \ T_b < 100^\circ\text{F}$	4

### 2.2.2 Inherent Safety Index (ISI)

Another notable safety index is the Inherent Safety Index (ISI) established by Heikkilä (1999) to serve as an extension of PIIS. A wider scope of safety parameters has been considered in ISI, in which their data must be readily available during preliminary process design phase. These parameters can also be classified into two groups, namely chemical inherent safety index and process inherent safety index. The former index includes sub-indexes for chemical reactivity, flammability, explosiveness, toxicity, and corrosiveness of chemical species present in the process; whereas the latter index contains sub-indexes for inventory, process temperature, process pressure, equipment safety, and safe process structure. The sub-indexes of equipment safety and safe process structure are not as straightforward as compared to other parameters as these two sub-indexes are evaluated based on accident statistics, layout data, case-based reasoning on a database of good and bad design cases, etc. The score domain differs from one sub-index to another, as the range of scores signifies the potential of impact of the specific sub-index to the plant safety. For instance, toxicity is deemed to have the most significant impact on plant safety, thus it has the largest score domain ranging from zero to six. On the other hand, corrosion is given the smallest score domain

(ranges from zero to two), as this particular hazard can be managed by an appropriate choice of construction materials. The total inherent safety index score of each process route is the summation of all sub-index values in both chemical and process inherent safety indexes. It is up to the process designer to introduce weighting factors to the sub-indexes when determining the total index score, if one considers that certain sub-indexes should be prioritised than others.

### **2.2.3 *i*-Safe**

Meanwhile, Palaniappan et al. (2002) has presented the *i*-Safe index which is mainly based on the sub-index values derived from PIIS and ISI to compare inherent safety of different chemical routes. Some additional factors are also taken into consideration when ranking process routes, namely worst chemical index, total chemical index and worst reaction index. The sub-indexes of flammability, toxicity, explosiveness, and NFPA reactivity rating are utilised to determine the chemical index, while the sub-indexes of temperature, pressure, yield, and heat of reaction assess the reaction index.

### **2.2.4 Other Inherent Safety Indexes**

Besides index-based approach, Gupta and Edwards (2003) presented a simple graphical method using the sub-indexes from PIIS to compare the inherent safety level of several chemical process routes. Meanwhile, Khan and Amyotte (2004) introduced the Integrated Inherent Safety Index (I2SI) which considers the process life cycle with economic assessment and hazard potential identification for each alternative. Sub-indexes involving hazard potential, inherent safety potential and add-on control equipment are applied. As PIIS, ISI and *i*-safe have treated the

chemicals present in a process route as individual components, where the chemical with the worst hazard (highest penalty score) is chosen to represent the chemical scores in these indexes, Leong and Shariff (2009) have thus proposed the process route index (PRI) to address the mentioned concerns. PRI considers all chemicals present as a mixture, and the explosiveness level is used as the measurement tool to quantify the inherent safety level for process route selection. Shariff and Leong (2009) have presented an inherent risk assessment in preliminary design stage for the development of process design. Its principal focus is to determine the inherent risk of an explosion through the application of 2-region F-N curve. Shariff and Zaini (2013) then extended this approach by proposing a 2-region risk matrix concept to carry out the risk evaluation for toxic release. Ahmad et al. (2014) developed a novel method known as the Numerical Descriptive Inherent Safety Technique (NuDIST) to assist in deciding the safest route in petrochemical industry. Recently, Ahmad et al. (2016) presented a graphical approach known as the Graphical Descriptive Technique for Inherent Safety Assessment (GRAND) to further examine several chemical process routes by identifying the root-cause of the hazards via visualisation method.

In conclusion, the inherent safety indexes as discussed in Section 2.2 are able to assist process designer to rapidly compare and rank several process routes manufacturing the same chemical during early design phase. Several key safety parameters that can result in major adverse impacts to plant safety are identified, and score domain for each parameter is constructed to reflect the hazard level of each process route. The information for chemical properties and process conditions of the routes must be readily available to apply the indexes. As process design progresses from early conceptual design phase to later phases, more

decisions will be made and it is not effective to modify these decisions later due to cost and time constraints. Hence, the inherent safety assessment approach is most effective when applied in early stage, where potential hazards of each process route can be identified early. By quantifying the hazard level of each route using the index based approach, the best process route with the least chemical and process-related hazards can be rapidly selected. This helps to minimise the capital and operating costs needed on managing plant safety, as less add-on safety equipment are required in the final design. Besides inherent safety, the concept of inherent occupational health has also emerged as a crucial decision-making factor in plant design. The following section discusses on the development of inherent occupational health over the years.

## **2.3 Inherent Occupational Health**

As for the aspect of health, some of the earlier established health indexes include Chemical Exposure Index (CEI) (Dow Chemical, 1994) and Toxicity Hazard Index (Tyler et al., 1996), in which both indexes assess the short-term and acute health hazard risks resulted from possible chemical release to on site employees within process plants and neighbouring communities. Meanwhile, Koller et al. (2000) developed the EHS method to classify and examine potential safety, health, and environmental (SHE) hazards during process development. However, the aspect of health only serves a minor role in EHS method and it only addresses the exposure-effect relationship of the workplace, which includes irritation and chronic toxicity. Later, Johnson (2001) established the Occupational Health Hazard Index (OHHI), which was the first index to evaluate occupational health hazard in design phase. However, one drawback of OHHI is that certain factors are over-evaluated since excessive data are required. In order to

improve the shortcoming of OHHI, Hassim and Edwards (2006) then developed the Process Route Healthiness Index (PRHI), which is further discussed in the following section.

### **2.3.1 Process Route Healthiness Index (PRHI)**

PRHI aims to evaluate the potential occupational-related hazards of new process plants by considering numerous factors that may bring about impact on human health in the workplace. The two key components deemed to have direct influence on the health hazard level are the chemical substances present and the amount of chemical released. In order to determine the PRHI of a process route, five indexes which contribute to PRHI are developed, namely Inherent Chemical and Process Hazard Index (ICPHI), Health Hazard Index (HHI), Material Harm Index (MHI), Worker Exposure Concentration (WEC) and Occupation Exposure Limit (OEL). ICPHI is used to determine the work activities, process conditions and material properties that are potentially harmful to human health. This is assessed based on the assignation of penalty score to the activities, process conditions and material properties. HHI is used to evaluate chemicals that have the inherent ability to cause typical occupational disease to the workers. The information to determine this particular index is taken from the Occupational Safety and Health Administration (OSHA), Health Code (HC) and Health Effects (HE), which indicate the main effects of exposure to each chemical substance. Meanwhile, MHI is used to rank the healthiness of all materials involved in each process stage via NFPA Ranking for Health. However, the application of PRHI is rather complicated since it involves a broad range of parameters to be assessed, while some information may not necessarily be available especially in the early design phase. An enhanced index known as the Inherent Occupational Health Index (IOHI) by Hassim and Hurme (2010a) is then introduced.

### 2.3.2 Inherent Occupational Health Index (IOHI)

IOHI as presented by Hassim and Hurme (2010a) is one of the well-established inherent health indexes. The principal objective is to evaluate health risks of process routes during process research and development phase that can potentially cause adverse health effects to the workers. The degree of health hazard is largely affected by the potential harm caused by the material and the potential for exposure to the material present in the route. There are two indexes in IOHI, namely Index for Physical and Process Hazards ( $I_{PPH}$ ) and Index for Health Hazards ( $I_{HH}$ ). The former index focuses on the likelihood of workers to be exposed to the chemicals, while the latter index concentrates on health impacts and risks due to the exposure. In order to assess both  $I_{PPH}$  and  $I_{HH}$  on a process route, several factors or parameters that might bring about significant contribution to the arising of adverse health impacts are selected. Mode of process ( $I_{PM}$ ), material phase ( $I_{MS}$ ), volatility ( $I_V$ ), pressure ( $I_P$ ), corrosiveness ( $I_C$ ), and temperature ( $I_T$ ) are the six sub-indexes selected for  $I_{PPH}$  while  $I_{HH}$  is comprised of the sub-indexes of exposure limit ( $I_{EL}$ ) and R-phrase ( $I_R$ ). The sub-indexes are penalised based on the basis of worst situation, which means that the maximum penalty received by the most hazardous chemical in a reaction step will represent the sub-index penalty score for that particular reaction step.  $I_{PPH}$  is calculated by the sum of all its six sub-indexes while  $I_{HH}$  is the summation of exposure limit and R-phrase sub-indexes. The total IOHI for each process route is the sum of  $I_{PPH}$  and  $I_{HH}$ .

All the inherent safety and occupational health indexes highlighted in Sections 2.2 and 2.3 applied the same approach to assess and quantify the inherent hazard level of a process route, which is the index-based approach. All parameters or sub-indexes are allocated scores to reflect the degree of potential hazard present in a route. The inherent hazard index

must be applicable during early stage in process design and development (Tyler, 1985). The approach can be easily mastered by any users with different background due to its simplicity. In addition, index method is relatively helpful in decision making since it is convenient and not too time-consuming (Gnoni and Bragatto, 2013).

### **2.3.3 Applications of Inherent Safety and Occupational Health Indexes**

Numerous works have applied the established inherent safety and occupational health indexes to assist with chemical process route selection. Srinivasan and Nhan (2008) presented the Inherent Benign-ness Indicator (IBI) to evaluate the inherent benign-ness of a process route based on 15 parameters linked to safety, health and environmental impact. The safety parameters are similar to the indexes selected by *i*-Safe except for toxicity, which is considered under occupational health. Some limitations of the inherent safety indexes that include subjective scaling and weighting of factors are addressed with the application of a statistical analysis-based procedure presented in IBI. Objective scaling approach is employed to normalise parameters to range from zero (non-hazardous) to one (most hazardous). A frequency distribution of heat of reaction, pressure and temperature for common reactions are applied to scale these three parameters. Ee et al. (2015) have developed a Combined Approach for Inherent Safety and Environmental (CAISEN) assessment that aims to identify potential environmental and safety hazards during early stages of process design. The environmental impact is evaluated by the life cycle assessment (LCA), while the process safety is cover by all sub-indexes provided in ISI. The ISI percentage is applied to indicate the degree of



hazards, which is a ratio of the total calculated ISI score for a specific process over the maximum ISI score that a process can obtain.

Serna et al. (2016) have conducted a multi-criteria decision analysis for the selection of chemical process route with an integrated sustainability approach. The three principal sustainability dimensions considered are economic, environmental and social aspects. The social dimension can be divided into safety and occupational health indicator groups, in which the scorings for the former indicator groups are adopted from the normalised scales proposed by Srinivasan and Nhan (2008). Warnasooriya and Gunasekera (2017) have proposed an integrated index known as the Inherent Chemical Process Route Index (ICPRI) to evaluate the possible toxicological effects on environment and occupational health, and the chemical and process-related safety impacts to the plant. The evaluation of occupational health follows the method proposed by Hassim and Hurme (2010b), which takes into account the workplace chemical concentration present in the plant and the chemical exposure limit of all chemicals to determine the Occupational health Hazard Index (OhHI). As for inherent safety, the parameters considered are inventory, flammability, explosiveness, reactivity, temperature and pressure, where the scores are based on PIIS, ISI and NFPA ratings.

Generally, the parameters applied in the indexes can be classified into two groups. The process-related parameters examine the inherent safety hazards and the potential for hazards exposure posed by the process itself e.g. the process operating conditions. Meanwhile, the chemical-related parameters are measured by the physical and chemical properties of the molecule. Since it is stated in Section 1.3 that the safety and health aspects should be included as design objectives in a CAMD problem, the

chemical parameters from the indexes can be integrated into the CAMD problem to evaluate the safety and health performance of the generated molecules.

## **2.4 Computer-Aided Molecular Design (CAMD)**

In CAMD, the principal goal is to identify the molecule or molecular structure from the given set of molecule building blocks and a specified set of target properties (Gani, 2004b). In the early problem formulation phase, the desirable target properties that meet the customer's needs are defined, along with the target range. The appropriate chemical building blocks used to synthesise the molecules are also selected. A database comprising of large quantity of chemical groups is applied to explore a wide range of conventional molecules or novel molecular structures that can serve as the potential molecular candidates. Structural feasibility constraints are imposed to ensure that chemical structural rules are not violated. Molecules that meet the target properties are examined in the performance analysis stage, in which molecules that display undesirable attributes are removed, with the assistance of solving algorithm (Heintz et al., 2014b). An optimisation technique is utilised to systematically develop and evaluate molecules until the optimal molecular candidates have been identified (Papadopoulos, et al., 2010).

In CAMD, the physical and chemical properties of the molecules are generally estimated using some types of fragment-based methodology, such as group contribution (GC) methods, topological indices (TI), and signature descriptors (SD). The following sections depicts on the development of GC methods over the years.

### **2.4.1 Group Contribution (GC) Methods**

Most chemical process and product design often involve the knowledge of a set of physical and thermodynamics properties of pure compounds and mixtures. However in many cases, the property data for the components of interest may not always be readily available. One way to obtain such information is by conducting experimental works, which generally require substantial amount of time and resources. Alternatively, one can utilise property prediction methods to rapidly estimate the list of properties with adequate reliability. One of the conventionally applied estimation techniques is the UNIFAC method, a group-contribution (GC) based prediction approach first proposed by Fredenslund et al. (1975). UNIFAC model is developed based on the basic concept of GC approach, where the interactions of each functional group present in the molecules are represented by group-interaction parameters. These parametric coefficients are then applied to estimate the activity coefficients in nonideal liquid mixtures.

Nowadays, GC methods are the most commonly used property prediction techniques in CAMD (Austin et al. 2016). Through this approach, it considers that a molecule is formed from a collection of molecular groups. For instance, an ethanol compound can be constructed from the groups of  $\text{CH}_3$ ,  $\text{CH}_2$  and  $\text{OH}$ . Each group carries a distinct contribution value to a specific property. In order to predict a particular property, the contributions of groups present in the compound are summed up to determine the property value of interest. Through GC method, the property of a component can be estimated swiftly without the necessity of considerable computation resources. Various studies have been carried on estimating the properties of pure compounds via GC methods. One of the commonly referenced GC method works was presented by Joback and Reid

(1987), in which it was assumed that there are no interactions between the groups. Thus, the property prediction only considers additive contributions among groups present in the molecule. In this work, GC methods for eleven pure component properties are developed, which include normal boiling point, melting point, critical temperature, critical pressure, critical volume, ideal gas heat of formation, ideal gas Gibbs energy of formation, heat of vaporisation at normal boiling point, heat of fusion, ideal gas heat capacity, and liquid dynamic viscosity. The group contribution definition can be represented by Equation (2.1), where  $P$  represents a specific property,  $C_i$  is the contribution of group of type- $i$  with  $N_i$  occurrences, and  $f$  represents certain function of multiplication between  $N_i$  and  $C_i$ .

$$P = f\left(\sum_i N_i C_i\right) \quad (2.1)$$

Meanwhile, Constantinou and Gani (1994) established a new property estimation approach where the prediction is conducted at two levels. The primary level contains contributions from first-order functional groups, while the secondary level comprises of a set of second-order groups having the first-order groups as building blocks. The proposed model is given by Equation (2.2), where  $C_i$  is the contribution of the first-order group of type- $i$  that occurs  $N_i$  times,  $D_j$  the contribution of the second-order group of type- $j$  that occurs  $M_j$  times,  $f(P)$  is a simple function of the property  $P$ , and  $x$  is a binary value allowing user to decide whether to include secondary level prediction. This method is capable of differentiating isomers and offers an improved accuracy and reliability of property estimation. Apart from several properties predicted by the Joback and Reid (1987) method, Constantinou and Gani (1994) also included GC model for estimating standard enthalpy of vaporisation at 298 K. With the

incorporation of second-order groups, all but one of the compared property models exhibit an average relative error (ARE) of less than 3%.

$$f(P) = \sum_i N_i C_i + x \sum_j M_j D_j \quad (2.2)$$

Constantinou et al. (1995) later extended the Constantinou and Gani (1994) method to include the prediction of acentric factor and liquid molar volume at 298 K. Marrero and Gani (2001) presented an enhanced group-contribution based prediction of pure compound properties that includes three levels estimation. The primary and secondary levels are similar to those developed by Constantinou and Gani (1994), while the tertiary level comprises of groups that cover extensive structural information about molecular fragments of components whose description is inadequate via the first and second level groups. This approach also provides the prediction of complex heterocyclic and large polyfunctional acyclic compounds. The GC model now takes the following general form:

$$f(P) = \sum_i N_i C_i + x \sum_j M_j D_j + z \sum_k O_k E_k \quad (2.3)$$

where  $E_k$  is the contribution of the third-order group of type- $k$  that has  $O_k$  occurrences in a component, while  $z$  is a binary value to allow user to decide whether to incorporate third-order groups. This method demonstrates higher estimation accuracy as most of the compared property models possess an ARE of lower than 3%. Meanwhile, the most conventionally applied GC method for acute toxicity was developed by Martin and Young (2001), whom correlated the acute toxicity (96-h  $LC_{50}$ ) to the fathead minnow for almost 400 organic chemical compounds. The model is formulated using multilinear regression and computational neural

networks (CNNs). Marrero and Gani (2002) employed the Marrero and Gani (2001) model to further develop methods for the prediction of octanol-water partition coefficient ( $K_{ow}$ ) and aqueous solubility ( $W_s$ ) at ambient temperature. The contributions for all groups are determined by linear regression analysis based on a vast data set of  $K_{ow}$  and  $W_s$  values. Meanwhile, Gani et al. (2005) addressed the limitation of some GC methods in which certain groups and/or their contributions to a specific property may not be available. A group contribution<sup>+</sup> (GC<sup>+</sup>) technique is established where the missing groups are generated and their contributions can be estimated via a set of zero-order and first-order connectivity indices (CI). This work also compares the results generated from the GC technique, CI-based approach, and the GC<sup>+</sup> (combined GC-CI) method. It is observed GC approach provides a better accuracy provided that the contributions of all groups in the compound are available. In the case where missing groups are present in the chemical, GC<sup>+</sup> approach is the better option. Conte et al. (2008) then proposed the GC<sup>+</sup> models for the prediction of surface tension and viscosity of pure organic components. Nannoolal et al. (2008) extended the normal boiling point GC model to allow the estimation of vapour pressure. The inputs needed for this new model are the molecular structure of the chemical and its normal boiling point to determine the vapour pressure of that chemical at different temperatures.

Hukkerikar et al. (2012b) modified and enhanced the model parameters for GC<sup>+</sup> models used for property estimation with the employment of covariance matrices to quantify uncertainties in the predicted property values. There are 18 pure compound properties examined in this work, which include new GC<sup>+</sup> models for entropy of vaporisation at normal boiling point, flash point, auto ignition temperature,

Hansen solubility parameters, and Hildebrand solubility parameter. Hukkerikar et al. (2012a) also introduced GC<sup>+</sup> property models to predict 22 environment-related properties of organic compounds, such as oral rat *LD*<sub>50</sub>, bioconcentration factor, permissible exposure limit, global warming potential, and etc. Recently, Frutiger et al. (2016a) employed the Marreno and Gani (2001) GC method to establish new GC models for flammability-related properties that comprise of lower and upper flammability limits, flash point, and auto ignition temperature.

In summary, GC approaches are simple tools to use as they only require the occurrence of groups present in a compound and their contributions to be input into simple function models to calculate the property values of interest. GC methods are particularly useful when employed in CAMD since they can easily represent a vast and diverse chemical space as the groups can be bonded in numerous ways to generate a great variation of different molecular structures. In addition, GC techniques can be easily transformed into the mathematical formulations of CAMD problems as the GC function models and count of groups can be integrated in the context of mathematical optimisation (Austin et al. 2016).

Besides GC method, topological indices (TI) and molecular signature descriptors (SD) are also commonly used in CAMD. TI is defined from the concept of molecular graph theory, where atoms and bonds forming a molecule are considered as vertices and edges respectively. The connections between atoms can be expressed in terms of numerous matrices, such as vertex adjacency matrix, edge adjacency matrix, incidence matrix, etc. (Raman and Maranas, 1998). Randic (1975) developed the molecular connectivity indices (CI) to describe the degree of branching of molecules with the use of first-order molecular connectivity

index. Kier and Hall (1976) then included higher orders of CI to account for heteroatoms. Meanwhile, Visco et al. (2002) and Faulon et al. (2003) proposed the molecular SD for encoding local neighbourhood of a molecule. This technique represents atoms present in a molecule with the extended valencies to a user-specified parameter known as the signature height. The height of the signature denotes the level to which the neighbourhood information is determined (Chemmanurgattuvalappil et al., 2010). An atom is designated as a root atom, and the atomic signature covers all atoms/bonds extending out to the predefined height, without backtracking. The SD approach is capable of differentiating between stereoisomers and has been employed in many applications, which include descriptor in several biological platforms (Visco and Chen, 2017).

Even though all GC, TI, and SD methods like many property estimation techniques promise high predictive accuracy and reliability, the issue of uncertainty do exist in the estimated property values, which may affect the generation of optimal molecules in CAMD and process design. The following section further discusses the works done to manage uncertainty mainly due to property models.

#### **2.4.2 Property Prediction Uncertainty**

Most CAMD problems have coupled with property prediction methods, such as GC methods, for property estimations. The reliability and accuracy of these prediction methods have significant impacts on the effectiveness of CAMD model in identifying the optimal molecule for a specific design problem. In the past three decades, numerous works have been conducted to study the effect of uncertainty in property prediction models. According to Larsen (1986), the design of chemical process requires various types of data such as physical, thermodynamic, and



transport property data for pure chemicals and mixtures to assist with material and energy balances and for designing equipment. These data are also essential to determine technical feasibility and economic potential, compare process alternatives, and perform process alteration to account for changing operating conditions. The accuracy of the property data is crucial since it may affect process design and cost. Larsen (1986) proposed the data quality methods to prioritise the available data sources, identify the data range of applicability, develop quality codes to measure reliability, and apply numerical estimates for data accuracy. Mathias and Klotz (1994) analysed the performance of several thermodynamic properties models by utilising them on applications comprising of van der Waals (vdW) equations of state (EOS).

Meanwhile, Whiting (1996) emphasised the importance of the accuracy of physical property data and prediction models for examining the risk of process failure and the acceptable range of safety-factor in process design. Monte Carlo simulations and probabilistic sensitivity analyses were conducted to assist in quantifying thermodynamics-induced process uncertainties. A comparison study on the results of applying process simulations using different models was considered to provide process design engineer with detailed uncertainty information. Maranas (1997) was the first to present a systematic methodology to evaluate the effect of property prediction uncertainty in optimal molecular design problems. Multivariate probability density distributions method is employed to explicitly quantify uncertainty by modelling the probability of different realisations of group contribution parameters. Yan et al. (2003) performed a reliability analysis through the comparison of four GC methods (Joback-Reid, Constantinou-Gani, Wilson-Jasperson, and Marrero-Pardillo) to assess their capability in predicting the critical temperatures of organic

compounds. They concluded that the Constantinou-Gani method performs better with alkanes and alkenes, while the user-friendly Wilson-Jasperson model demonstrates compatible prediction results with all organic and inorganic compounds.

Dong et al. (2005) have reviewed 194 papers since 1940 that have studied the expression of uncertainty for thermodynamic property measurements specifically to calculate critical temperature for pure chemical compounds. The evolution and progression on the nature and extent of estimations of uncertainty have been emphasised. Even though the reporting of uncertainty data has shown enhancement over the years, comprehensive uncertainty analyses remained uncommon then, specifically with respect to the incorporation of contributions resulted from sample impurities. Hajipour and Satyro (2011) re-evaluated the Riazi and Daubert (1980) and Lee and Kesler (1980) correlations to estimate the critical temperature, critical pressure, and acentric factor of pure hydrocarbons in the range of  $C_5$  to  $C_{36}$ . The former have introduced the improved correlations which took into account the uncertainty of normal boiling point and specific gravity to estimate the uncertainties on critical properties and acentric factors. Hukkerikar et al. (2012b) presented the enhanced model parameters for group-contribution<sup>+</sup> ( $GC^+$ ) models to assist with the estimation of pure chemical properties. Parameters estimation and uncertainty analysis were considered on the  $GC^+$  models with the use of the maximum-likelihood estimation method to present property predictions with higher reliability, along with an estimate of prediction uncertainties of the predicted property values. Ng et al. (2015a) addressed the uncertainties arised from the application of property prediction models in a CAMD problem through the introduction of property robustness. This property robustness signifies the accuracy of the prediction models, which

is expressed in terms of standard deviation of the models that serves as a measurement of average variation between the actual experimental values and the prediction values.

Recently, Frutiger et al. (2016b) established a methodology to deal with the numerical and statistical concerns when developing GC models, specifically for the estimation of heat of combustion in this study. Uncertainty analysis on the model parameter estimation and property prediction was considered to improve the new heat of combustion GC method with better estimation performance statistics. Comprehensive uncertainty analysis with 95% confidence interval of the prediction model was performed to ensure that the estimated properties of interest are of high accuracy. For most GC methods developed in literature, their capability of estimating the properties of pure compounds are demonstrated by their corresponding statistical performance indicators. Some of the indicators include coefficient of determination ( $R^2$ ), standard deviation (SDE), average absolute error (AAE), and average relative error (ARE). In this work, GC methods and other empirical correlations are used to estimate the molecular properties that are evaluated in the inherent safety and health sub-indexes. Thus, the accuracy and reliability of the estimation models highly affect the accuracy of the allocation of sub-index scores to the molecules. This has resulted in the issue of uncertainty resulted from the application prediction models. To address this concern, uncertainty from property prediction is managed on the safety and health sub-indexes. The statistical performance indicators provided by the models identify, which are then adopted to determine the uncertain region at all property boundaries present in the sub-indexes. The sub-index scores in this region are adjusted and enhanced accordingly to better reflect the molecular inherent hazard level under property prediction uncertainty.

### 2.4.3 CAMD Applications

CAMD techniques have been adopted extensively in many chemical industries, such as the design of chemical-based products, solvents, active ingredients, pharmaceutical drugs, polymers, refrigerants, lubricants, extractants, catalysts, ionic liquids, biodiesel additives, and more. One of the early CAMD works was conducted by Gani and Brignole (1983), which synthesised potential solvents for the separation of aromatic and paraffinic hydrocarbons on the basis of UNIFAC predictions through enumeration method. Brignole et al. (1986) then extended and improved the problem formulation for the synthesis, assessment and screening of solvents. UNIFAC groups are first characterised based on their respective combinatorial properties, then the appropriate groups are selected depending on the nature of problem. Molecular structures are identified by solving the partitioned combinatorial problem subjected to a list of feasibility criteria. Joback (1989) presented a molecule design procedure with six primary steps: problem formulation, target transformation, group selection, molecular enumeration, molecular screening, and final evaluation. Case studies on the design of refrigerants, polymers, solvents, and drugs were considered by applying the proposed methodology. Gani et al. (1991) have proposed a structured CAMD algorithm with four steps: pre-specify molecular groups and target properties, synthesise a feasible set of optimal molecular structures, estimate the properties and screen the set of molecules, and design the final molecular compound. This work also considered the selection and ability of the computational methods to generate optimal solutions with high predicted performance. Klein and Wu (1992) incorporated optimisation techniques in computer-aided mixture design problems to determine the chemical identity and composition of solvent mixtures that are optimised with respect to targeted properties. The objective function of the model is expressed by a cost function, which

serves to rank the solvent mixtures, while some thermodynamic properties act as constraints to be complied. A two-loop optimisation algorithm known as the successive regression and linear programming (SRLP) is established to solve the model with the presence of non-linear constraints. Odele and Macchietto (1993) pioneered the utilisation of mixed-integer nonlinear programming (MINLP) in CAMD problems which requires that the number of each molecular group type present in the molecule must be in the form of integer values. This formulation allows nonlinearity in objective and constraint functions comprising of both continuous and discrete decision variables. The methodology is illustrated with solvent design case studies for liquid extraction and multicomponent gas absorption. Venkatasubramanian et al. (1994) first introduced the application of genetic algorithms in CAMD to address the limitations encountered by some of the earlier CAMD works. Genetic algorithms provide systematic stochastic search for enhanced solutions through sampling areas of the parameter space that offer higher potential for good solutions. Besides, genetic algorithms are capable of directly integrating advanced level of chemical knowledge and reasoning strategies for a higher efficiency search. Churi and Achenie (1996) presented a novel mathematical programming model to generate chemical compounds that can achieve the targeted performance characteristics. It applies discrete values which are able to provide an almost complete structural and connectivity information of the molecule. Another advantage is that property prediction methods with high accuracy can be utilised to their full extent in the proposed model. A case study on refrigerant design for the replacement of Freon 12 has been considered to demonstrate the application of the formulation. Maranas (1996) introduced a framework that reformulates a class of optimal molecular design problems which incorporated the nonlinear structure-property functionalities into the corresponding mixed-integer linear

programming (MILP) problems. The proposed framework is capable of eliminating the caveat of convergence to suboptimal molecular designs and enhancing the possibility of generating novel and superior design solutions. The methodology is illustrated with a case study on polymers design, while addressing two design objectives that include property matching and property optimisation. The former functions to minimise the maximum scaled deviation of design properties from some target values, while the latter serves to minimise or maximise a single property subject to the pre-specified lower and upper bounds of the remaining properties. Constantinou et al. (1996) established a general framework for computer-aided product design (CAPD) that adopts group contribution method to synthesis a wide range of compounds with various degrees of complexity and size. A new group contribution based approach for property prediction with improved accuracy is adopted to enhance the validity of generated solutions in achieving the targeted performance. The framework is then demonstrated on the design of solvents, polymers, azeotropes and miscible mixtures, and minimum cost solvent for paint and coating applications. Instead of applying mathematical programming and combinatorial techniques that have been proposed in many previous CAMD works, Ourique and Silva Telles (1998) introduced an alternative which employs molecular graphs for the representation of chemical compounds and utilises simulated annealing algorithm in the search of target compounds. These two principal features of the methodology are applied in solvent design for extraction of n-butanol and refrigerant design for the replacement of chlorofluorocarbons (CFCs) in heat pumps. Raman and Maranas (1998) first integrated topological indices as structural descriptors for correlating properties within a MILP optimisation framework. The topological indices provide full molecular interconnectivity information and exhibit higher accuracy of property correlations than simple group contribution methods.

Two examples that include the design of alkanes attaining the targeted properties and the substituent selection for optimal fungicidal and insecticidal properties have been considered using the presented framework. Harper et al. (1999) extended the work done by Constantinou et al. (1996) to couple molecular modelling techniques with a traditional CAMD method. This work enhances the compound selection features and improves the generation of molecular structures and property prediction associating to molecular modelling. The extended CAMD uses a multilevel procedure with each level carrying out its own generation and test steps. The results from each level are served as the input for the subsequent level so that the combinatorial problem can be solved efficiently.

Marcoulaki and Kokossis (2000) have proposed a computer-aided technique for the synthesis of molecules with optimal properties. The approach integrates stochastic optimisation and group-contribution methods to develop chemicals of desirable performance. Three different case studies have been illustrated using this approach, namely the design of refrigerants to substitute Freon-12, the design of a heating medium, and the design of a solvent for the liquid-liquid extraction of ethanol from water. As most mathematical programming based CAMD problems are often formulated as MINLP with significantly large amount of linear constraints as compared to nonlinear/nonconvex constraints, most search space variables in the form of binary (integer) and continuous variables are usually presented in the nonlinear terms. As a result, Ostrovsky et al. (2002) implemented a branch-and-bound algorithm (BB) for branching on a group of linear branching functions that are linearly dependent on the search variables in a reduced dimension. A case example to design cleaning agents in lithographic printing has been considered to demonstrate the proposed strategy. Karunanithi et al. (2005) introduced a

computer-aided molecular/mixture design framework that is capable to develop optimal solvents and solvent mixtures that are optimised with respect to structural, property and process constraints. As the CAMD problem is formulated as a MINLP, it is solved using a proposed decomposition-based solution method. Two case studies involving the design of extractant to separate acetic acid from water in a liquid-liquid extraction process and the design of pharmaceutical compound in a real industrial problem are illustrated with the presented methodology. Yamamoto and Tochigi (2008) have carried out the molecular design of foaming agents to search for hydrofluoroethers as potential alternatives using neural network (NN) method, which is a data-processing approach containing input, hidden and output layers. The thermophysical properties needed for the design of a foaming agents, such as the enthalpy of vaporisation, surface tension and normal boiling point, have been estimated via NN method. Several generated compounds are selected based on thermal conductivity as determined by multiple regression. Folić et al. (2008) introduced an extended hybrid experimental/computer-aided technique to design solvents used in reactions. First, a reaction model is developed from several reaction rate measurements, then the CAMD problem is formulated to synthesise the optimal solution with respect to maximising the reaction rate constant. Satyanarayana et al. (2009) have employed CAMD approach to generate the polymer repeat unit structures which comply with the required constraints. A systematic framework integrating recently developed group contribution plus ( $GC^+$ ) models and an extended CAMD approach to include design of polymer repeat units has been proposed. Through this integration, a vast amount of polymer structures can be taken into consideration in polymer design problems. Meanwhile, the design of liquid formulations is often conducted by experiment-based trial-and-error technique, which is time-intensive and



involves large amount of resources. Conte et al. (2011) proposed the integrated experiment-modelling approach, which is a hybrid of computer-aided model-based techniques and heuristic based experimental testing, for the development of liquid formulated products. The approach is tested on the design of paint formulation and insect repellent lotion, in which a set of feasible product candidates are synthesised for each design problem and the chemical candidates are validated through a course of pre-specified activities (work-flow). Chemmangattuvalappil and Eden (2013) applied the molecular signature descriptors for the design of molecules in a reverse problem formulation framework. The signature descriptors offer to trace the alterations in molecular groups present in a molecule resulting from different kinds of chemical reactions. Hence, the changes in the molecular structure caused by reactions are expressed as a function of the property. The problem is then solved to synthesis molecular structures that achieve the property targets in the process design step. Meanwhile, the search for the suitable binary working fluid mixtures to be used in organic Rankine cycle (ORC) has been conducted by Papadopoulos et al. (2013) using CAMD approach. The potential molecular candidates serving as the first compound of the binaries are first identified. The next step is to select the best matching molecules and subsequently determine the optimum mixture concentration to attain optimum mixture performance targets. Mattei et al. (2013) has applied mixture property models and CAMD techniques for the design of emulsion-based chemical products. Their methodology comprises of seven sequential hierarchical steps and a conceptual case study using this methodology to develop a personal detergent is demonstrated. Stavrou et al. (2014) utilised and extended the approach of continuous molecular targeting-computer-aided molecular design (CoMT-CAMD) which integrated both solvent and process optimisation of a precombustion CO<sub>2</sub>-capture system with physical absorption. The process topology involved in

this study comprises of all major process operations of a conventional CO<sub>2</sub>-capture system, and the performance is assessed with a single economic objective function. Quantitative structure property relationship (QSPR) models for the ideal gas heat capacity and molar mass of the optimised solvent have been constructed via the pure component parameters of the PC-SAFT equation of state. Meanwhile, Cignitti et al. (2015) introduced a CAMD framework for the development of pure, mixed and blended chemical based products. The molecular design formulation is transformed into a MINLP, which is then sequentially solved using a decomposed optimisation technique. Gebreslassie and Diwekar (2015) presented a novel approach to solve for the design of optimal solvent molecule to extract acetic acid from process waste stream based on an efficient ant colony optimisation (EACO) algorithm. The problem is solved using MINLP model by maximising the performance of the solvent with respect to structural feasibility, property and process constraints. Ng et al. (2015b) discussed the developments, current obstacles and future opportunities in the chemical product design area. Some significant challenges discussed include the necessity to develop new property prediction models, the computational complexity in CAMD, and the need to introduce a sustainable chemical product design framework with the growing focus on greener future. Dev et al. (2016) presented a multi-objective CAMD algorithm to formulate both reactants and products in which their respective targeted properties are optimised. As the objective functions of the work include structure-dependent properties, both group contributions and/or topological indices are adapted into the algorithm. Recently, Chong et al. (2017) presented a methodology for the development of optimal ionic liquid solvent for CO<sub>2</sub> capture in a bio-energy system. In addition to identifying the optimal solvent, the approach is able to determine the optimal operating conditions which offer the best carbon capture

performance. Scheffczyk et al. (2017) incorporated quantum mechanical information into CAMD to assist with property prediction using COSMO-RS method ('Conductor-like Screening Model for Realistic Solvation'). As a result, an optimisation-based framework known as COSMO-CAMD is introduced to generate new solvents. This work has also considered case studies to design potential solvents in liquid-liquid extraction of phenol and hydroxymethylfurfural (HMF) from water. Abedin et al. (2017) developed a computational technique to design water compatible visible light photosensitisers that could enhance the photo-polymerization of the hydrophilic-rich phase in dental adhesive resin. A model building set is applied to formulate the QSPRs for properties associated to the photo-polymerization reaction of the adhesive monomers and hydrophilicity of the photosensitiser through the utilisation of connectivity indices. The QSPRs are then integrated into the problem framework to synthesise photosensitiser molecules with optimised properties.

As previously mentioned in Sections 1.3 and 1.4.2, the aspects of safety, health and environment have been considered as design constraints in many CAMD problems. Numerous earlier works were prompted by the search of less harmful substitutes to replace chlorofluorocarbons (CFCs) as refrigerants, since it was discovered that the use of CFCs can result in the depletion of ozone layer. Previously, Duvedi and Achenie (1996) have utilised CAMD mathematical programming for the design of environmentally-friendly refrigerants. The augmented-penalty/outer-approximation approach is applied to solve the MINLP problem for the generation of optimal compounds that are optimised with respect to a performance index, which is represented in terms of explicit target properties. The main focus of the environmental constraint used is the ozone-depletion potential (ODP), which is predicted through correlations

linking with the number of halogenated groups found in the component. Meanwhile, a three-step approach introduced by Pistikopoulos and Stefanis (1998) also aimed to synthesise solvents with lessened environmental impacts. The framework involves the identification of agent-based process operations, determination of a set of solvent blends complying with the processing and environmental constraints, and screening of an optimal solvent blend with regards to global plant-wide basis constraints. Environmental impact assessment methods such as the post-release environmental damage were used as the environmental criteria in solvent selection. Trade-offs between the cost and environmental impact were also examined in this work. Meanwhile, a design problem comprising of many nonlinear correlations can result in a nonconvex formulation, which may lead the problem to be trapped in a local solution. In order to tackle this issue, Sinha et al. (1999) proposed a global optimisation model to solve a solvent design problem to guarantee the yield of a global solution. An environmental constraint on the octanol-water partition coefficient ( $\log K_{ow}$ ) is imposed in their framework to generate environmentally benign solvents for the applications of surface cleaning in printing industry. Buxton et al. (1999) presented the procedure for the selection of optimal solvent blends to be used for nonreactive and multicomponent absorption processes. This work is an extension of the solvent design technique presented by Pistikopoulos and Stefanis (1998). The solvent alternatives are evaluated with respect to process operational, environmental and economic performance. The environmental impact of the blends was assessed by the lethal concentration,  $LC_{50}$  constraint to denote their respective toxicity to the environment. Chavali et al. (2004) employed connectivity indices to carry out physical property predictions of a transition metal catalyst, which include electronegativity, density and toxicity. The toxicity used for the conducting environmental evaluated is in the form of  $LC_{50}$  (lethal dose in

air for 50% of exposed animals), which has been correlated based on the database from the Syracuse Research Corporation. The connectivity indices are then incorporated in the optimisation approach for the redesign of environmentally-benign homogeneous catalysts with respect to the three mentioned target properties. As for safety aspect, Karunanithi et al. (2006) presented a CAMD algorithm in a MINLP model for the design and selection of solvents and/or anti-solvents for solution crystallisation. Besides satisfying the solvent property requirements such as solubility and crystal morphology, safety constraints in terms of flammability and toxicity are also imposed to prevent the generation of hazardous solvents. Patel et al. (2010) adopted the concept of inherent safety design (ISD) to develop solvent processes, where the principle of 'substitution' is practised by utilising CAMD technique as a tool for the selection of inherently safer solvents. Similar safety constraints in the forms of flash point and  $\log LC_{50}$  toxicity are also considered to screen out flammable and toxic solvents. Meanwhile, consequence models and regulatory guidelines provided by EPA Risk Management Program are employed to incorporate safety consideration in process simulation to ensure an inherently safer process design. Papadokonstantakis et al. (2015) integrated several criteria such as thermodynamic performance, environmental impact, hazard evaluation, and economic analysis with predictive molecular-based approaches, CAMD, and process modelling to develop a solvent-based post-combustion CO<sub>2</sub> capture system. Two sustainability indices have been considered in this work, namely the life cycle analysis metrics that deal with comprehensive environmental impact for standard process operation, and hazards metrics that examine the harm potential in accidental situations. In this work, the EHS method by Koller et al. (2000) and the Hazard Identification and Ranking System (HIRA) by Khan and Abbasi (1998) are applied as the tools to assess hazards. Palma-Flores et al. (2016) carried out

simultaneous molecular and process design for the application of ORC to recover low-temperature waste heat. The evaluation of the generated working fluids was conducted using the flammability index from Kondo et al. (2001) and the toxicity index by Toxicity Estimation Software Tool (TEST) to assure that the working fluids attained the desired sustainability characteristics. Gerbaud et al. (2017) extended the methodology of Heintz et al. (2014a) that involves all stakeholders within the chemical enterprise to search for novel commodity chemicals using a combined computer-aided organic synthesis (CAOD) and CAMD technique. In order to ensure that the new products can comply with the increasingly stricter environmental, health and safety specifications, property constraints for flash point, acute toxicity ( $LC_{50}$ ), octanol-water partition coefficient ( $\log K_{ow}$ ), and bioconcentration factor are imposed. Two case studies involving the substitution of aprotic highly dipolar solvents with itaconic acid derivatives and the design of biolubricants from vegetable oils have been carried out. Many of these CAMD problems involve multiple design objectives, in which several target properties are selected as the objective functions to be optimised. In order to solve a multi-objective design problem, some of the developed multi-objective optimisation tools can be employed, and they are discussed in Section 2.4.4.

#### **2.4.4 Multi-Objective Optimisation**

In many CAMD problems, the generation of optimal molecules must meet many criteria, such as performance characteristics, economic potential, environmental impact, and more. A molecule with high performance attributes may be harmful to the ecosystem, such as the early use of CFCs in refrigeration applications. These problems can be represented as multi-objective optimisation problems, in which the aim is to evaluate the trade-off between different criteria with respect to a set of

optimal solutions (Ghosh and Chakraborty, 2015). Generally, the methods applied to solve such problems can be grouped into two categories, namely ideal procedure and preference-based procedure. The former procedure first determines several trade-off optimal solutions with a wide range of values for the objective functions. Subsequently, the best trade-off solution is selected using higher-level information. Meanwhile, the ideal-based multi-objective optimisation procedure solves the problem with a composite objective function to form the weighted sum of the objectives, where the weight assigned to a specific objective is dependent on its preference factor. This preference-based procedure is much simpler to be applied, but it is relatively subjective compared to the ideal procedure (Deb, 2001). One the most regularly used classical preference-based approach is known as the weighted sum method (Fishburn, 1967). This approach is user-friendly as all objectives are multiplied with their respective user supplied weight, and they are then summed up to scalarise into a single objective. However, the method is rather subjective as the values assigned to the weights are based on the user's judgement, in which they are assigned in a manner that larger weight represents higher importance to the problem. Besides, most data and parameters available for multi-objective problems are generally indefinite and uncertain in nature (Singh and Yadav, 2015). Thus, the application of traditional techniques may not be efficient to tackle these problems with inherent imprecision. An alternative to solve multi-objective optimisation problem is the fuzzy optimisation algorithm. Zadeh (1965) first proposed the fuzzy set theory to represent and manipulate fuzzy data that was not precise. A fuzzy set can be defined as a group of elements with a continuum of degrees of membership, ranging from zero to one. The proposed method is known to be a capable tool to solve a multi-objective problem under fuzzy environment. Bellman and Zadeh (1970) then included the concept of fuzzy

theory in a decision-making problem. Each objective function is first converted into a fuzzy goal, then all fuzzy goals together with the fuzzy constraints are expressed in terms of fuzzy sets in the space of alternatives. A fuzzy solution can be depicted as an intersection of the specified fuzzy goals and constraints. An optimal solution is then illustrated by a point in the space of alternatives where the membership function of the specific fuzzy solution attains its maximum value. Meanwhile, Zimmermann (1978) extended the Bellman-Zadeh principle to present the fuzzy linear programming method, in which a fuzzy decision set of solutions is expressed through the aggregation of all fuzzy inequalities by utilising the 'min' aggregation approach. The optimal solution of the linear program can be then obtained by the one fuzzy decision for which its minimum aggregated function is the maximal (Dubey and Mehra, 2014).

Over the years, max-min aggregation method has been widely applied in numerous research works. Amid et al. (2011) integrated a weighted max-min fuzzy model in a supply chain problem to decide on the selection of supplier. Since many input information for such problem are not known precisely, the proposed model offers to manage the vagueness of input data and different weights of criteria in this problem. The weights of criteria assigned in this problem are decided by the application of analytic hierarchy process (AHP). Tay et al. (2011) employed fuzzy optimisation for the development of a sustainable integrated biorefinery with two conflicting objective functions, namely economic and environmental performance. Pareto optimality analysis is then performed to examine results generated via max-min aggregation approach. Meanwhile, Ng et al. (2014b) introduced a systematic framework to construct a bioenergy-based industrial symbiosis (BBIS) that includes biorefineries, existing milling facilities and combined heat and power (CHP)



plants operated by different owners. Since each owner has distinct profit-oriented goals, this problem is managed with the concept of industrial symbiosis (IS) to deal with their cooperation. The utilisation of a novel disjunctive fuzzy optimisation technique facilitates in achieving the optimum network configuration with maximum economic performance which also satisfies the respective interests of all participating owners. Ng et al. (2015a) also adopted max-min aggregation method in fuzzy optimisation to manage the trade-off between property superiority and property robustness in a CAMD problem. The former demonstrates the optimality of product property; while the latter measures the accuracy of property prediction models to estimate molecular properties. The proposed methodology offers the synthesis of optimal molecules under property prediction uncertainty. Wan et al. (2016) presented a fuzzy multi-footprint optimisation (FMFO) approach for Malaysian sago industry to synthesis a sustainable value chain. The proposed approach simultaneously takes into account carbon footprint, water footprint, workplace footprint, and economic potential as objectives. Max-min aggregation in fuzzy optimisation is then employed to balance the conflicting objectives in order to select the optimal sustainable sago value chain. In this research, fuzzy optimisation can help in optimising two main objective functions in this work, namely the target performance of the molecule and the inherent safety and health level of the molecule itself. In order to determine the hazard level of molecules, inherent safety and health sub-indexes are employed to provide index scores to the molecules. As discussed in Section 1.4.4, one way in improve the quantification of overall molecular hazard is by introducing weight factors to the sub-indexes. One method to determine such weights is through the application of analytic hierarchy process (AHP), which is further discussed in the following section.

## **2.5 Analytic Hierarchy Process (AHP)**

AHP is described by Saaty (1994) as a theory of relative measurement that is now widely used for many decision-making problems. AHP is a systematic procedure that offers the depiction of a decision-making problem in hierarchical structure. The problem is decomposed into smaller constituent parts or elements, and a series of pairwise comparison judgments among elements are carried out with respect to a common criterion or feature. A scale of values is applied in the pairwise comparisons to express the relative strength or intensity of impact of one element to another. Subsequently, the priority of each element is derived based on the input judgements to select the preferred alternative. Hence, AHP is a powerful technique that is appropriate for the management of multi-objective, multi-criterion, and multi-actor decisions with any number of alternatives. The key advantage of using AHP is its ability to examine and minimise the inconsistency of different expert judgments (Aminbakhsh et al., 2013).

AHP has gained much popularity in many decision-making applications due to its simplicity, flexibility, intuitive appeal, and its capability to blend quantitative and qualitative specifications in a single decision framework (Ramanathan, 2002). Odynocki (1979) applied AHP as the principal tool of examination to identify the National Health Insurance proposal among three plans that has the highest potential to enhance the health care system. The conceptual model of the "ideal" health care system is formulated with all its elements weighed through pairwise comparisons with respect to their impact to the system. Saaty and Gholamnezhad (1982) assessed different approaches for the safe disposal of high-level nuclear waste. Five disposal strategies have been compared and prioritised with regards to a list of tangible and intangible criteria. Arbel (1987)

presented an AHP based technique for the identification of the best market entry alternative to venture into a new technological market. The alternatives are compared with a list of assessment criteria that include managerial control, financial details, sales effectiveness, technical infrastructure, and etc. Schniederjans and Wilson (1991) introduced a new hybrid of the AHP and zero-one linear programming (ZOLP) methodologies to assist with the selection of information system projects and the assignment of resources to accomplish them. AHP helps to prioritise the set of information system projects subjected to the appropriate criteria of the organisation, while the prioritisation data is subsequently applied as a ranking strategy within the structure of a ZOLP model. The latter model also takes into account the resource availability constraints encountered by the organisation for the appropriate selection of projects. Akash et al. (1999) conducted a comparison between several electricity power generation alternatives in Jordan, which include fossil fuel power plants, nuclear, solar, wind, and hydro-power energies, through the application of AHP. Cost-to-benefit analysis is carried out with the construction of separate hierarchies for benefits and costs.

Handfield et al. (2002) demonstrated the utilisation of AHP serving as a decision support tool to assist purchasing managers for comprehending the trade-offs between different environmental dimensions. AHP offers to analyse the relative significance of several environmental attributes and evaluate the relative performance of different supplies with respect to these attributes. Promentilla et al. (2006) proposed an assessment technique for the prioritisation and selection of countermeasures at the planning phase of site remediation. The analytic network process (ANP) supermatrix approach, which is an extension of AHP to manage decision structure with higher complexity, is utilised to input

decision maker's value judgment in order to compute the relative preference of the remedial alternatives. Four decision models obtained from the generalised hiernet are analysed to illustrate the role of hierarchic functional dependence, inner dependence, and feedback cycle in influencing the computation of priority weights. Aminbakhsh et al. (2013) introduced AHP in the safety risk evaluation procedure to prioritise the safety risk items in construction project. Besides, AHP also functions as a tool to examine and minimise inconsistencies of safety risk severities provided by the expert judgments. The cost of safety (COS) model is also adopted for the planning of rational budget and establishing practical goals without compromising safety. Recently, Li et al. (2016) established a task analytic hierarchy approach based on AHP to assess multiple-criteria task and decision-making in nuclear safety. Several criteria and sub-criteria of task are adapted into the method, which can be categorised into objectivity, person and security. This strategy is capable of selecting optimal scheme with minimised hazard and improved efficiency of task. Jagtap and Bewoor (2017) demonstrated an application of AHP to determine the major equipment in a thermal power plant. The ranking of equipment is subjected to four main criteria, namely effect on failure of equipment on power generation, environment and safety, frequency of failure, and maintenance cost. Kluczek (2017) illustrated an overall multi-attribute method for sustainability evaluation in manufacturing processes. AHP is applied to rank activity areas for manufacturing process with respect to four dimensions of sustainability, which cover the technical, environmental, economic, and social aspects. The capability of the presented framework is demonstrated through a case study in the production of heating devices.

In conclusion, AHP is considered a useful tool to solve a multi-criterion or multi-attribute decision-making problem with three steps: decompose the problem into a hierarchical structure, pairwise comparisons among the elements or alternatives to a common goal, and synthesise priority for each element to identify the preferred solution. In this work, as the overall hazard level of a molecule is determined by sum of the safety and health sub-indexes, it is necessary to introduce weights to the sub-indexes to represent the level of contribution of each sub-index to the overall impact. A higher significance sub-index is prioritised by allocating a higher weight to it. Hence, AHP can be adequately applied to identify the appropriate weight factors. The application of AHP in weight determination is demonstrated in Chapter 7.

## **2.6 Summary**

In order to integrate the concept of inherent safety and occupational health in CAMD, the background theory of the development of inherent safety and occupational health is discussed in Sections 2.2 and 2.3. Inherent safety design strives to reduce the amount of hazardous chemical substances used in the process and to operate processes with milder conditions. The principles of inherent safety design are best applied in the early design phase as potential hazards can be identified and decisions can be made earlier to eliminate or minimise the impact of hazards. Numerous inherent safety indexes have been developed to examine the safety level of different process routes, such as PIIS, ISI, and *i*-Safe. These indexes employed safety-related parameters/sub-indexes, in which their respective data must be available in early design phase, to conduct safety evaluation of the chemical process routes. The overall hazard level can be determined by the summation of all sub-index scores involved in the assessment. As for inherent occupational health, the focus is to reduce the occupational

health-related hazards from chemical processes to the employees. The two key inherent occupational health indexes are PRHI and IOHI, which utilised health-related sub-indexes to evaluate health risks. This index-based approach is a convenient and simple tool to rapidly quantify and rank different chemical process routes with respect to their safety and health level. Thus, it can also be easily adopted in the CAMD programming to examine the molecular safety and health performance.

As the concept of inherent safety and occupational health is adopted to replace the use of dangerous chemicals with less harmful ones, CAMD is an adequate tool to search for the suitable chemical candidates with promising targeted performance. CAMD is capable of identifying the molecular structure that satisfies the specified target properties from a given list of chemical building blocks. Some group-contribution based property prediction models have been commonly incorporated in CAMD programming to assist with the estimation of properties depending on the molecular structure. CAMD technique has also been widely applied in many chemical applications, such as the design of solvents, polymers, refrigerants, and etc. The consideration of safety and health aspects in many CAMD works have also been discussed, while the shortcomings of the existing works are also highlighted. The research gap will serve as the main motivation of this research work, which is to simultaneously consider both safety and health aspects along with the property functionalities to function as design objectives to be optimised in the CAMD framework. As there are multiple criteria to be taken into account, multi-objective optimisation technique is employed to optimise all design objectives simultaneously. The proposed max-min aggregation fuzzy-based optimisation is able to carry out the trade-off between the objectives. Hence, it can optimise the two main objectives in this research, which are

the target functionality performance and the safety and health attributes of the molecules.

Meanwhile, the background study of AHP is also discussed, which is a tool to handle multi-objective or multi-attribute decision-making problem. It decomposes the problem into smaller constituents in a hierarchical structure, then pairwise comparisons are considered among elements with respect to a common goal or criterion. Numerical scales are used and the priorities of all elements are determined to select the preferred solution. The application of AHP is proposed in the fourth scope to assist the determination of weight factors which will be assigned to the sub-indexes. A larger impact sub-index is prioritised by allocating it with higher weight to ensure that it offers a greater contribution to the overall safety and health level of the molecule. The next chapter illustrates the research methodology of the four proposed scopes.

## CHAPTER 3

### RESEARCH METHODOLOGY

The overall methodology for the integration of safety and health aspects into CAMD framework is presented, which includes analysing molecular performance with inherent safety and health sub-indexes, incorporating safety and health as design objectives along with targeted property functionalities, managing the allocation of safety and health sub-index scores under property prediction uncertainty, and enhancing the sub-index scores by score smoothing and introducing weight factors. The methodology is divided into six research stages. The research methodology proceeds with the following stages: early research stage, research scope 1, research scope 2, research scope 3, and ends with research scope 4.

#### 3.1 Early Research Stage

The early stage of this research work is to carry out the background study on CAMD, and to determine the research gaps in the current state of CAMD. The following steps are considered in this stage:

- i. Literature study on existing CAMD works and future direction.
- ii. Identify research gaps in current CAMD works.
- iii. Generate ideas and identify research direction.
- iv. Establish scopes of research.

#### 3.2 Research Scope 1

Evaluate the safety and health aspects of optimal molecules using inherent safety and health indexes



The aim of this research scope is to develop a CAMD model to generate optimal molecules that meet the desired target properties as specified by user. The generated set of molecules then undergo a performance analysis stage to evaluate their safety and health performance with the application of inherent safety and occupational health sub-indexes. The following steps are the brief procedure in this scope:

- i. Define molecular design goal of the problem.
- ii. Identify target properties of the problem.
- iii. Select target properties to serve as design objectives to be optimised.
- iv. Search for appropriate property prediction models to estimate property values.
- v. Select appropriate molecular building blocks to synthesis molecules.
- vi. Introduce structural constraints to ensure that only feasible molecules can be formed.
- vii. Apply fuzzy optimisation to solve the multi-objective optimisation problem to generate optimal molecular structures.
- viii. Select appropriate safety and health sub-indexes to serve as tools to assess the performance of the generated optimal molecules.
- ix. Assign sub-index scores to the optimal molecules to represent their inherent hazard level.
- x. Rank the molecules according to their corresponding total index score.

### **3.3 Research Scope 2**

Integrate inherent safety and health aspects into chemical product design framework

The principal aim of this scope is to consider both safety and health aspects along with the target property functionalities as design objectives in the CAMD model. The proposed framework is able to generate optimal molecules with high performance and favourable safety and health attributes. The following steps are employed in this stage:

- i. Follow steps (i) to (iii) of research scope 1 as given in Section 3.2.
- ii. Include the aspects of safety and health as design objectives along with the target functional properties.
- iii. Select appropriate safety and health sub-indexes to serve as tools to assess the performance of the potential molecules.
- iv. Search for appropriate property prediction models to estimate target properties and properties that are applied in the safety and health sub-indexes.
- v. Introduce disjunctive programming to convert property values into their corresponding sub-index scores.
- vi. Assign sub-index scores to molecules and determine the total index score of each molecule which represents the overall hazard level.
- vii. Follow steps (v) to (vii) of research scope 1 as given in Section 3.2.

### **3.4 Research Scope 3**

#### Manage uncertainty on the application of property prediction in safety and health sub-indexes

The primary aim of this research scope is to manage property prediction uncertainty in the safety and health sub-indexes. As the accuracy of sub-index scores depends on the reliability of the prediction models, the magnitude of uncertainty exhibited by each model is integrated into the sub-indexes so that the scores can be revised to better

reflect the hazard level of the molecules under uncertainty condition. The following steps are the key procedure in this scope:

- i. Follow steps (i) to (iv) of research scope 2 as given in Section 3.3.
- ii. Manage uncertainty resulted from the application of property prediction models to estimate properties evaluated in safety and health sub-indexes.
- iii. Identify standard deviation given by each of the property prediction model applied.
- iv. Locate every property boundary in the sub-indexes, which is the point where the sub-index score switches from one value to another.
- v. Determine the uncertain range around each property boundary with the standard deviation of the respective property.
- vi. Enhance the scores in the uncertain range with the use of a linear slope to better reflect the hazard level.
- vii. Follow steps (v) to (vii) of research scope 2 as given in Section 3.3.

### **3.5 Research Scope 4**

Improve the measurement of safety and health by introducing weight factors for sub-index prioritisation and smoothing sub-index scores

The main goal of this scope is to address the discontinuity of scores at property boundary when applying the sub-indexes. As the safety and health-related properties in the sub-indexes are divided into several sub-ranges, the property boundary separating two adjacent sub-ranges is the point in which the penalty score switches abruptly from one value to another. Thus, the scores in the boundary range are smoothed to ensure continuous transition of scores. Besides, different weight factors are also introduced to the sub-indexes so that larger impact sub-indexes are given

larger weights to allow great contributions to the overall safety and health level of molecules. The following steps are applied in this stage:

- i. Follow steps (i) to (iv) of research scope 2 as given in Section 3.3.
- ii. Locate every property boundary in the sub-indexes, which is the point where the sub-index score switches from one value to another.
- iii. Identify the range around each property boundary in which the sub-index scores will be smoothened.
- iv. Smoothen the sub-index scores around each property boundary range to ensure that the scores are continuous at any point.
- v. Introduce disjunctive programming to convert property values into their corresponding sub-index scores.
- vi. Apply analytic hierarchy process (AHP) method to determine weight factors that will be assigned to sub-indexes.
- vii. Assign weight factor to each sub-index depending on the severity of score; a sub-index with higher score is prioritised and given higher weight factor.
- viii. Determine the total index score of each molecule by summing up the multiplications between the scores and their respective weights.
- ix. Follow steps (v) to (vii) of research scope 1 as given in Section 3.2.

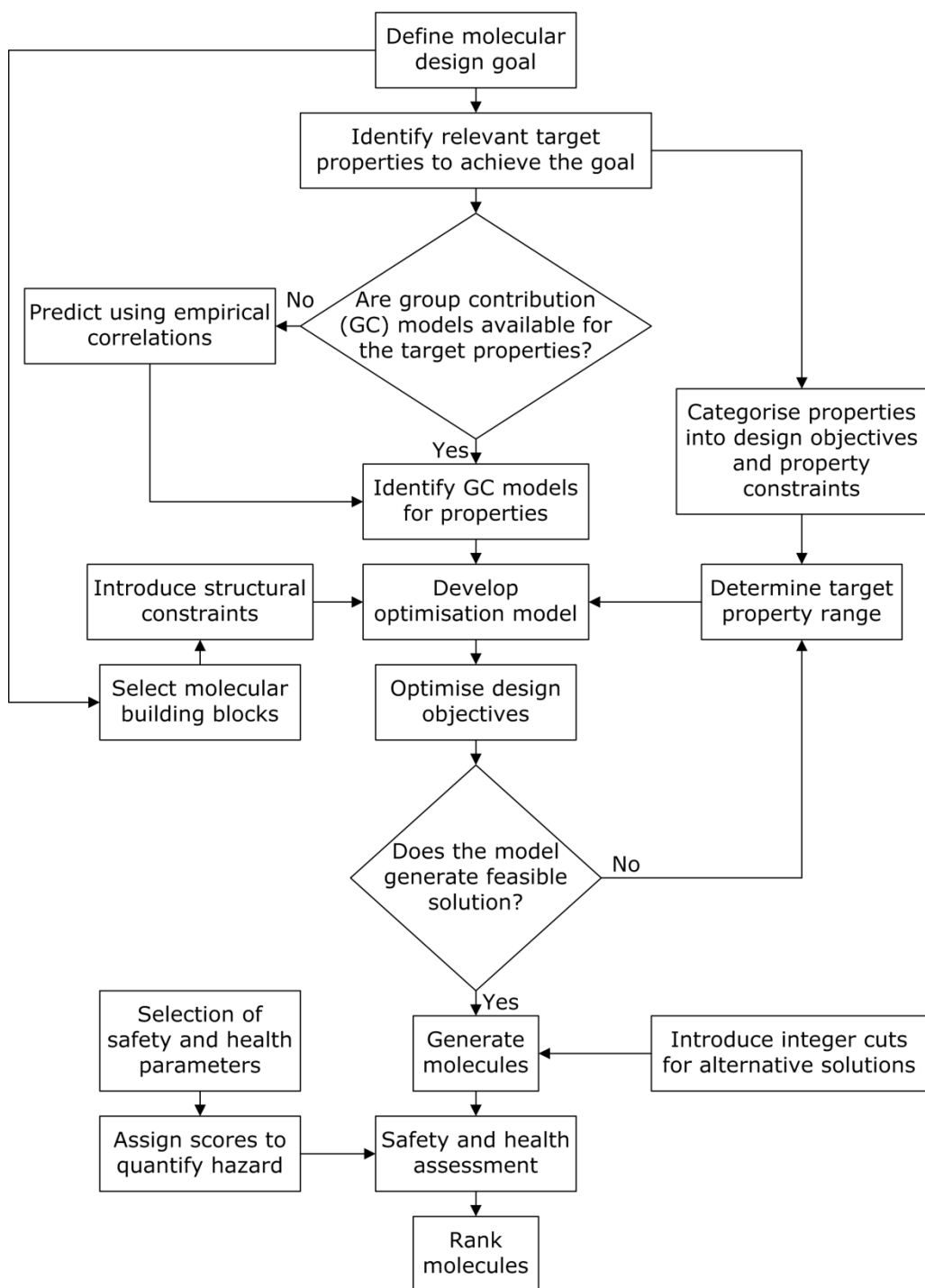
## 3.6 Methodology Overview

Figure 3.1 presents the flow diagram of the methodology applied in Research Scope 1, while Figure 3.2 is the methodology proposed for Research Scope 2. It is noted that both figures differ from one another as in Figure 3.1 (scope 1), safety and health aspects are only evaluated in the performance analysis stage once the molecules are generated. Meanwhile, safety and health aspects along with the target properties are considered

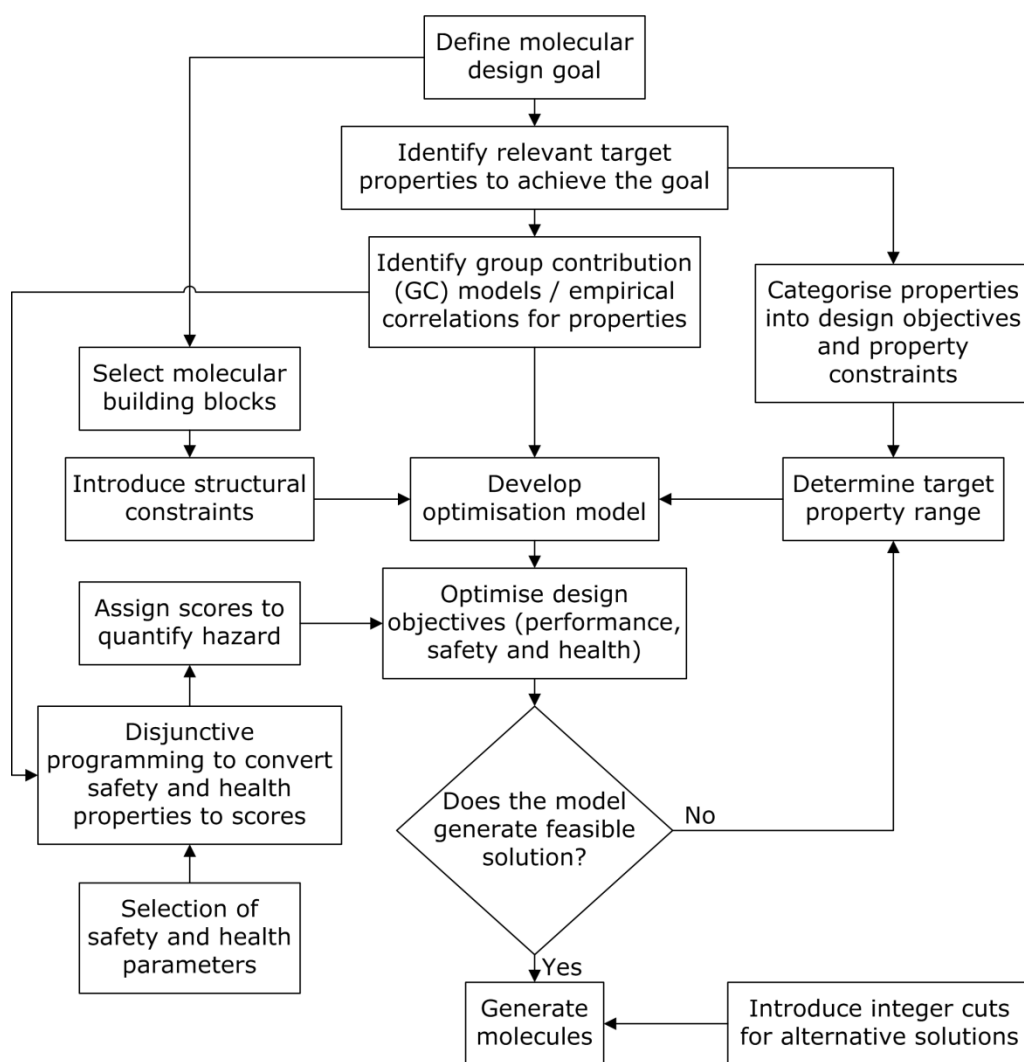
as design objectives in Figure 3.2 (scope 2). Figure 3.3 illustrates the methodologies of Research Scopes 3 and 4, which are the extensions of the procedure carried out in Figure 3.2 (scope 2). The additional steps in Figure 3.3 as compared to Figure 3.2 are represented by step boxes with thicker outline. The one step with single asterisk (\*) is only considered when carrying out the methodology in Research Scope 3, while the two steps with double asterisks (\*\*) are only adopted in the methodology of Research Scope 4.

### **3.7 Summary**

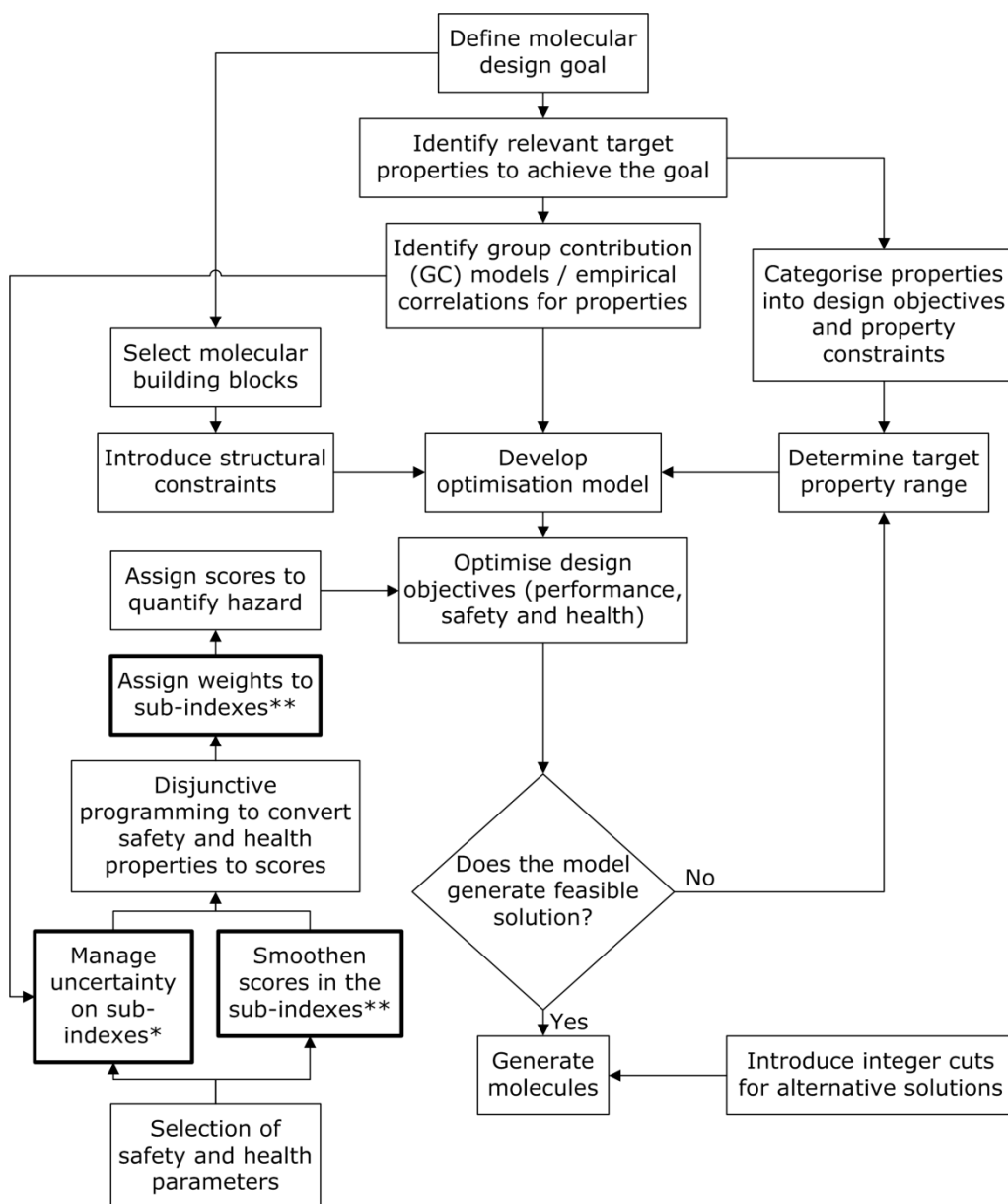
Chapter 3 presents the brief methodology of the four research scopes proposed in this research. The methodology of Research Scope 1 uses safety and health aspects as assessment tool in the performance analysis stage to screen and rank molecules with better safety and health features. The methodology of Research Scope 2 considers both aspects of safety and health as design objectives. The methodologies of Research Scopes 3 and 4 are the extensions of Scope 2. Scope 3 addresses the uncertainty resulted from property predictions in sub-indexes, while Scope 4 improves the safety and health measurement of molecules by smoothing the scores and allocating weight to each sub-index to differentiate the impact level among sub-indexes. The detailed methodologies of the four scopes are illustrated in the following Chapters 4 to 7.



**Figure 3.1:** Flow diagram of overall methodology in Research Scope 1



**Figure 3.2:** Flow diagram of overall methodology in Research Scope 2



**Figure 3.3:** Flow diagram of overall methodologies in Research Scopes 3 and 4



## **CHAPTER 4**

# **A NOVEL CHEMICAL PRODUCT DESIGN FRAMEWORK WITH THE INTEGRATION OF SAFETY AND HEALTH ASPECTS**

### **4.1 Introduction**

This chapter presents the first scope of the research, which is the utilisation of CAMD techniques to design an optimal molecule that meets the desired target properties. Since the CAMD problem involves multiple property functionalities, fuzzy optimisation is employed to deal with the trade-off between conflicting objectives to obtain optimal solutions. The generated set of optimal molecules is then assessed in the performance analysis stage with respect to the safety and health aspects. Inherent safety and occupational health indexes are used as the assessment tools to evaluate the safety and health performance of each molecule. These molecules are assessed by several safety and health-related sub-indexes, where penalty score representing the degree of hazard is allocated to each sub-index. The total index score obtained by each molecule is determined to quantify the inherent safety and health level of the molecule. The molecules are then ranked according to their safety and health index scores to highlight molecules with better performance. A case study on solvent design for gas sweetening process is performed to illustrate the application of the proposed methodology.

## 4.2 Problem Statement

In most CAMD problems, the selected design objectives to be optimised are usually represented by physical and thermodynamic properties of the molecule. After optimisation is carried out, the generated molecules are optimal with respect to the targeted functional properties. However, these molecules with high functionality are not guaranteed with favourable safety and health performance. Though safety and health aspects are evaluated during performance analysis phase, there is a lack of a systematic assessment methodology to integrate these aspects during the design stage. In this chapter, the existing inherent safety and health indexes, which have long been developed to evaluate the inherent hazard level of a chemical process route, are adapted in a CAMD problem to assess the inherent safety and health level of the generated molecules. This is to ensure that the selected molecule achieves the targeted functionality, and at the same time, does not cause major adverse impacts to human. The safety and health evaluation of the molecules is conducted during the performance analysis stage after the optimal molecules are generated. There are several specific problems to be addressed, which are stated as follows:

1. As CAMD problem involves the need to optimise multiple design objective properties which may be conflicting in nature, a formulation known as fuzzy optimisation can address the trade-off between these objectives and simultaneously optimise them.
2. Since most of the existing inherent safety and health indexes are developed for process route selection, not all factors or parameters considered in these indexes can be applied in the problem. The selection of appropriate safety and health-related sub-indexes has

to be carried out in order to evaluate the safety and health performance of the molecules.

3. These sub-indexes are usually represented as physical and chemical properties. In order to determine the property values of the molecules, appropriate property prediction models have to be identified. If there are no property prediction models available for a particular property, other empirical correlations of that property are identified.

## **4.3 Methodology**

The main objective of this scope is to establish a systematic framework that incorporates the aspects of safety and health as molecular evaluation tools in a CAMD problem. CAMD approach is first utilised to generate a list of optimal molecules that meet the functional target properties of the design problem. The generated molecules are then evaluated based on their safety and health performance with the application of inherent safety and occupational health indexes. This is to ensure that an inherently safer and healthier molecule that meets the desired target properties is selected as the final solution. This proposed design includes problem formulation, model development, molecular design, optimisation model, and performance analysis.

### **4.3.1 Problem Formulation**

In this stage, the needs of a chemical product are identified by defining the product characteristics and specifications to determine its functionality and physical behaviour. These specifications can be translated in terms of target properties. All selected target properties are then expressed with respect of property range, where the property value must

fall within the predefined range for the molecule to function and behave in the desired manner. This specification range can be illustrated with two inequality expressions bounded by lower bound ( $v_p^L$ ) and upper bound ( $v_p^U$ ) of the range as shown by Equation (4.1):

$$v_p^L \leq V_p \leq v_p^U \quad \forall p \in P \quad (4.1)$$

where  $p$  represents the target property, while  $V_p$  represents the target property value. The desired properties to be optimised are selected as the design objectives which the molecules must achieve in order to serve its function. The remaining properties will serve as property constraints for the molecules to fulfil.

### 4.3.2 Model Development

In this stage, all properties that are considered as the target properties have to be calculated through property prediction models. The most notable approach coupled in most CAMD problem is the GC method, which is able to estimate the physical and chemical properties of a molecule based on its structure. The general form of a GC model is shown in Equation (2.3). Meanwhile, properties with no GC method available can be estimated using correlations and empirical relationships.

### 4.3.3 Molecular Design

In this stage, the appropriate molecular groups acting as potential building blocks are chosen. For instance, if a CAMD problem requires the synthesis of alcohol-based molecules, then hydroxyl group (OH) must be selected. Next, structural constraints are specified and implemented in order to eliminate combination of infeasible solution. In order to ensure

that a single molecular structure is generated, the summation for the number of occurrences for all selected groups must be greater than zero:

$$\sum_{i=1}^{G_T} N_i > 0 \quad (4.2)$$

where  $N_i$  is the number of occurrences of group  $i$  while  $G_T$  is the total number of groups needed to form the molecules. Besides, in order to ensure that a molecule does not contain free bonds, the octet rule of structural feasibility is applied (Odele and Macchietto, 1993):

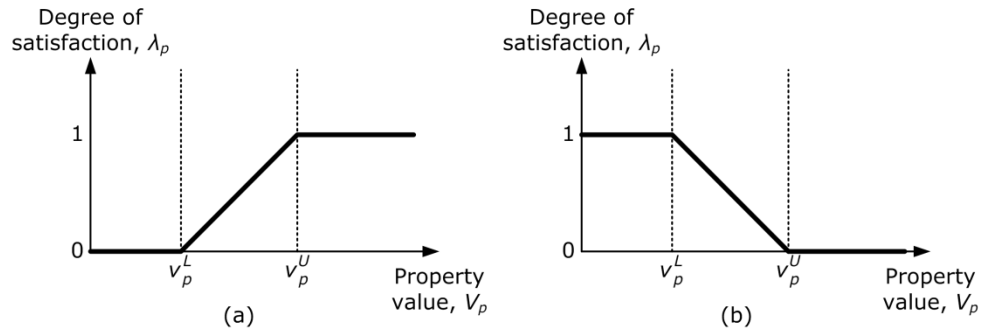
$$\sum_{i=1}^{G_T} N_i (2 - v_i) = 2g \quad (4.3)$$

where  $v_i$  is the valence of group  $i$  and  $g$  is 1, 0, -1 or -2 for acyclic, monocyclic, bicyclic and tricyclic compounds respectively.

#### 4.3.4 Optimisation Model

In this stage, the target property models are first transformed into their property operators. The property operators are represented by the linear combinations of the number of occurrence for molecular group of type- $i$  and its corresponding contribution. The lower and upper bounds of the property operators are then determined. This is further illustrated in Section 4.4.4.1. Based on the target properties identified in Section 4.3.1, some of them are selected as the design objectives in which the selected properties are optimised. The remaining target properties will act as property constraints to be fulfilled. Since multiple objectives are involved, the CAMD problem is now a multi-objective optimisation problem. However, some design objectives may potentially be conflicting in nature.

Fuzzy optimisation algorithm can be adopted as a decision making tool to deal with the trade-off between these conflicting objectives. In order to apply fuzzy optimisation algorithm, a degree of satisfaction,  $\lambda_p$  has to be introduced to each design objective property.



**Figure 4.1:** The degree of satisfaction,  $\lambda_p$  curve for design objective to be maximised (a) or minimised (b)

Figure 4.1 illustrates the degree of satisfaction curve for design objective to be maximised or minimised. As shown in Figure 4.1(a), the goal is to maximise  $V_p$ . Hence any  $V_p$  above  $v_p^U$  has a  $\lambda_p$  value of one, which indicates that the objective is fully satisfied. On the contrary, any  $V_p$  below  $v_p^L$  has a  $\lambda_p$  value of zero, which depicts that the objective is not satisfied. In between  $v_p^L$  and  $v_p^U$ , the  $\lambda_p$  value is represented by a linear function. The reverse mechanism is shown in Figure 4.1(b) where the goal is to minimise  $V_p$ . From Figure 4.1,  $\lambda_p$  can then be expressed as a linear membership function bounded by  $v_p^L$  and  $v_p^U$  as shown in Equations (4.4) and (4.5). Equation (4.4) is used for a design objective to be maximised while Equation (4.5) is used for a design objective to be minimised.

$$\lambda_p = \begin{cases} 0 & V_p \leq v_p^L \\ \frac{V_p - v_p^L}{v_p^U - v_p^L} & v_p^L \leq V_p \leq v_p^U \\ 1 & V_p \geq v_p^U \end{cases} \quad \forall p \in P \quad (4.4)$$

$$\lambda_p = \begin{cases} 0 & V_p \geq v_p^U \\ \frac{v_p^U - V_p}{v_p^U - v_p^L} & v_p^L \leq V_p \leq v_p^U \\ 1 & V_p \leq v_p^L \end{cases} \quad \forall p \in P \quad (4.5)$$

$$0 \leq \lambda_p \leq 1 \quad \forall p \in P \quad (4.6)$$

$\lambda_p$  is a continuous variable representing the level of satisfaction, which ranges from 0 to 1. In order for all  $\lambda_p$  to reach high level of satisfaction, all  $\lambda_p$  have to be maximised or be close to 1. Max-min aggregation method developed by Zimmermann (1978) can be employed in this work to maximise the least satisfied degree of satisfaction. This is to ensure that all values of  $\lambda_p$  will be satisfied partially to at least the degree of  $\lambda$ . A modified max-min aggregation technique is applied to ensure the that solution generated by the optimisation model achieves Pareto optimality (Javadian et al., 2009). The overall objective function is shown by Equation (4.7), subjected to constraint of Equation (4.8):

$$\max \lambda + \frac{1}{M} \sum_{p \in P} \lambda_p \quad (4.7)$$

$$\lambda_p \geq \lambda \quad \forall p \in P \quad (4.8)$$

where  $M$  is an arbitrarily large number. The huge value of  $M$  ensures that the second term in Equation (4.7) does not have significant effect on the final numerical value of  $\lambda$ . As the second term will assume a non-zero

value, this forces the individual  $\lambda_p$  to be maximised, which ensures Pareto optimality (Yu et al., 2016; Promentilla et al., 2017).

In order to generate multiple optimal solutions, integer cuts have been applied. Integer cuts introduce additional constraints into the problem formulation which prevent the formation of the optimal solutions that have already been identified. This approach assists the designer in finding alternatives other than the optimal solution. For instance, when the optimisation model has generated an optimal molecular structure, an additional constraint is then introduced to the model. The constraint ensures that the molecular structure of this particular optimal solution will not be formed again in the search of the next alternative. The integer cut constraint is formulated as follows:

$$\sum_{i=1}^{G_T} |N_i - N_i^*| > 0 \quad (4.9)$$

where  $N_i^*$  represents the number of molecular group of type  $i$  of the optimal solution(s) that has/have already been generated by the optimisation model. For example, a particular CAMD problem considers  $\text{CH}_3$ ,  $\text{CH}_2$  and  $\text{OH}$  groups as the molecular building blocks to formulate the molecules. If the first optimal solution generated by the model consists of one  $\text{CH}_3$  group, one  $\text{CH}_2$  group and one  $\text{OH}$  group (equivalent to an ethanol compound), the integer cut constraint added for next model iteration is given by:

$$|N_{\text{CH}_3} - 1| + |N_{\text{CH}_2} - 1| + |N_{\text{OH}} - 1| > 0 \quad (4.10)$$



where  $N_{\text{CH}_3}$ ,  $N_{\text{CH}_2}$  and  $N_{\text{OH}}$  represent the number of  $\text{CH}_3$ ,  $\text{CH}_2$  and  $\text{OH}$  groups present in the molecule respectively. With the addition of constraint (4.10), the next model iteration will not generate the exact same combination of molecular groups (or ethanol compound) as the next optimal solution.

### 4.3.5 Performance Analysis

**Table 4.1:** Parameters evaluated in inherent safety indexes

Parameters	PIIS	ISI	<i>i</i> -Safe
Heat of reaction		✓	✓
Heat of side reaction		✓	
Chemical interaction		✓	
Reactivity rating			✓
Flammability	✓	✓	✓
Explosiveness	✓	✓	✓
Toxicity	✓	✓	✓
Corrosiveness		✓	
Inventory	✓	✓	
Yield	✓		✓
Temperature	✓	✓	✓
Pressure	✓	✓	✓
Type of equipment		✓	
Process structure		✓	

The generated molecules from the previous stage then undergo the performance analysis stage to assess their safety and health performance. First, the selection of safety and health sub-indexes has to be carried out. The selection of the inherent safety and health indexes are made based on

the existing indexes that have already been well developed. For safety indexes, PIIS (Edwards and Lawrence, 1993), ISI (Heikkilä, 1999) and *i*-Safe (Palaniappan et al., 2002) are considered. Table 4.1 shows a summary of the parameters that are considered in their indexes.

However, only some of these parameters can be applied in this work (product design) as these indexes were not originally meant for product design; instead they are developed for selecting the 'best' chemical process route to synthesise the desired product. Hence, parameters that are related to process are selected in this work. The parameters that are related to the chemical properties are heat of reaction, heat of side reaction, chemical interaction, reactivity, flammability, explosiveness, toxicity, and corrosiveness. In a CAMD problem, it is easier to apply parameters that can be directly linked to the properties which can be estimated through property prediction methods. Parameters in which the index scores are assigned based on non-numerical descriptions may not be easily included in the mathematical optimisation model. For instance, the corrosiveness sub-index from ISI is based on the basis of the required construction material, such as carbon steel, stainless steel and special materials. Expert judgments are needed for the selection of appropriate construction material, as there are no direct prediction methods that can be utilised to decide on such selection. Therefore, two parameters are chosen from the safety indexes, namely flammability ( $I_{FL}$ ) and explosiveness ( $I_{EX}$ ). The former sub-index can be evaluated using flash point ( $F_p$ ) and normal boiling point ( $T_b$ ), while the latter sub-index can be measured by upper and lower explosion limits ( $UEL$  and  $LEL$  respectively). All these properties can be easily estimated through property prediction models. Toxicity exposure is not chosen as one of the safety sub-indexes to avoid repetition as it has already been included in one of the health indexes

(Hassim and Hurme, 2010a). The index scores for explosiveness sub-index are taken from ISI while the score for flammability sub-index is taken from NFPA flammability rating (National Fire Protection Association, 2007). Even though PIIS also offers explosiveness index scores, the maximum score assigned is ten, which is relatively high compared to the maximum index score given by ISI and NFPA. For consistency purpose, explosiveness index scores from ISI are applied. On the other hand, the reason for applying the NFPA flammability rating is that it is one of the standard systems that is commonly applied to classify the hazards of the materials, as it is frequently used in the material safety data sheet (MSDS). The penalty scores for the two safety sub-indexes are shown in Tables 4.2 and 4.3.

**Table 4.2:** Flammability ( $I_{FL}$ ) sub-index (National Fire Association Protection, 2007)

Parameter	Score Information	Penalty Score
Flammability, $I_{FL}$	Nonflammable	0
	$F_p \geq 93.4^\circ\text{C}$	1
	$F_p < 93.4^\circ\text{C}$	2
	$F_p < 37.8^\circ\text{C}$	3
	$F_p < 22.8^\circ\text{C} \ \& \ T_b < 37.8^\circ\text{C}$	4

**Table 4.3:** Explosiveness ( $I_{EX}$ ) sub-index (ISI) (Heikkilä, 1999)

Parameter	Score Information	Penalty Score
Explosiveness, $I_{EX}$	Non explosive	0
$S = (UEL-LEL) \text{ vol}\%$	$0 \leq S < 20$	1
	$20 \leq S < 45$	2
	$45 \leq S < 70$	3
	$70 \leq S \leq 100$	4

**Table 4.4:** Parameters evaluated in the inherent health indexes

Parameters	PRHI	IOHI
Mode of process	✓	✓
Material phase	✓	✓
Volatility		✓
Pressure	✓	✓
Corrosiveness	✓	✓
Temperature	✓	✓
Viscosity	✓	
Ability to precipitate	✓	
Density difference	✓	
Volume changes	✓	
Solubility	✓	
Exposure limit		✓
R-phrase		✓
Transport	✓	
Venting or flaring	✓	
Maintenance works	✓	

As for health indexes, the two indexes considered are PRHI (Hassim and Edwards, 2006) and IOHI (Hassim and Hurme, 2010a). As shown in Table 4.4, many of the parameters evaluated are mainly focussed on process aspects. Similar to the selection of safety parameters, only the chemical-properties parameters are chosen for molecular assessment, which include viscosity ( $I_\eta$ ) from the PRHI, as well as material phase ( $I_{MS}$ ), volatility ( $I_V$ ), and exposure limit ( $I_{EL}$ ) from the IOHI. For these four chosen health sub-indexes, the evaluation is carried out at 25°C and 1 atm as these are usually the conditions in which the workers handle the materials in the plant.  $T_b$  and normal melting point ( $T_m$ ) are used to determine  $I_{MS}$ ,

while  $T_b$  alone is used to examine  $I_V$ . According to Hassim and Hurme (2010a), the exposure limit is with the basis of 8-h daily exposure time and it illustrates a chronic type of toxicity, which is represented in terms of permissible exposure limit ( $PEL$ ). Besides chronic-typed toxicity, the acute type of toxicity is equally important and it should also be covered in this work. Hence, another parameter named acute health hazard ( $I_{AH}$ ) is also included in which the scoring for this sub-index is based on the NFPA health hazard rating (National Fire Protection Association, 2007). From the NFPA health hazards, the potential of a material to cause injury due to contact with or entry into the body via inhalation, skin contact, eye contact, or ingestion is addressed. These can be measured using  $LC_{50, inhalation}$  for acute inhalation toxicity,  $LD_{50, dermal}$  for acute dermal toxicity, and  $LD_{50}$  for acute oral toxicity. Since the group contribution model for  $LD_{50}$  (acute oral toxicity) is available, this particular property is used in this sub-index to carry out the acute toxicity measurement. The penalty scores for the five selected health sub-indexes are shown in Tables 4.5 to 4.9. The total penalty score of a molecule ( $I_{SHI}$ ) is the summation of all the sub-index scores assigned to it, which is shown in Equation (4.11). A molecule with lower total penalty score is desired as it represents an inherently safer and healthier molecule.

$$I_{SHI} = I_{FL} + I_{EX} + I_{\eta} + I_{MS} + I_V + I_{EL} + I_{AH} \quad (4.11)$$

**Table 4.5:** Viscosity ( $I_{\eta}$ ) sub-index (PRHI) (Hassim and Edwards, 2006)

Parameter	Score Information	Penalty Score
Viscosity, $I_{\eta}$	Low ( $0.1 \text{ cp} \leq \eta < 1 \text{ cp}$ )	1
	Medium ( $1 \text{ cp} \leq \eta < 10 \text{ cp}$ )	2
	High ( $10 \text{ cp} \leq \eta \leq 100 \text{ cp}$ )	3

**Table 4.6:** Material phase ( $I_{MS}$ ) sub-index (IOHI) (Hassim and Hurme, 2010)

Parameter	Score Information	Penalty Score
Material phase, $I_{MS}$	Gas	1
	Liquid	2
	Solid	3

**Table 4.7:** Volatility ( $I_V$ ) sub-index (IOHI) (Hassim and Hurme, 2010)

Parameter	Score Information	Penalty Score
Volatility, $I_V$	<i>Liquid and gas</i>	
	Very low volatility ( $T_b > 150^\circ\text{C}$ )	0
	Low ( $150^\circ\text{C} \geq T_b > 50^\circ\text{C}$ )	1
	Medium ( $50^\circ\text{C} \geq T_b > 0^\circ\text{C}$ )	2
	High ( $T_b \leq 0^\circ\text{C}$ )	3

**Table 4.8:** Exposure limit ( $I_{EL}$ ) sub-index (IOHI) (Hassim and Hurme, 2010)

Parameter	Score Information	Penalty Score
Exposure limit, $I_{EL}$	<i>Vapour (ppm)</i>	
	$PEL > 1000$	0
	$PEL \leq 1000$	1
	$PEL \leq 100$	2
	$PEL \leq 10$	3
	$PEL \leq 1$	4

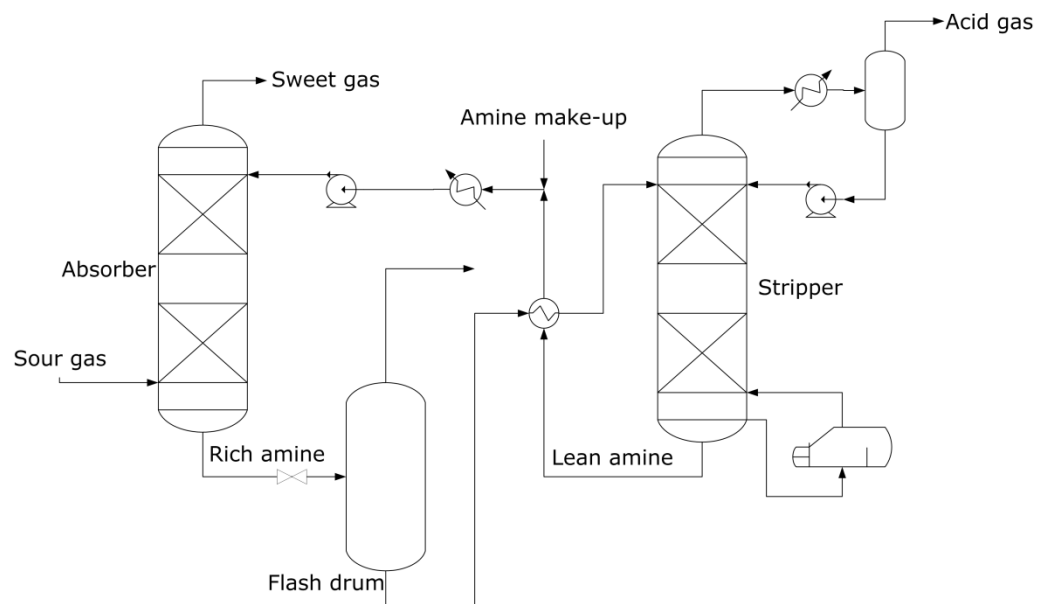
**Table 4.9:** Acute health hazard ( $I_{AH}$ ) sub-index (National Fire Association Protection, 2007)

Parameter	Score Information	Penalty Score
Acute health hazard, $I_{AH}$	<i>Oral rat <math>LD_{50}</math> (mg/kg)</i>	
	$LD_{50} > 2000$	0
	$500 < LD_{50} \leq 2000$	1
	$50 < LD_{50} \leq 500$	2
	$5 < LD_{50} \leq 50$	3
	$LD_{50} \leq 5$	4

#### 4.4 Case Study: Solvent Design for Gas Sweetening Process

Over the years, many anthropogenic activities have resulted in the release of huge quantity of carbon dioxide ( $CO_2$ ) into the atmosphere. Fossil fuel combustion can be regarded as one of the major contributions leading to the increase in  $CO_2$  emissions. This can bring about adverse impact to the environment, i.e. global warming. Nowadays, chemical industries have played a major role in promoting sustainable development. The removal of  $CO_2$  from fossil fuel combustion is one critical issue that must be addressed. The most extensively used technology for the removal of  $CO_2$  is the chemical absorptions with the utilisation alkanolamines as solvents (Muhammad and GadelHak, 2015). Figure 4.2 shows a schematic diagram of a gas sweetening process using amines as absorbent. The sour gas enters at the bottom of the absorber column while the lean amine solution enters the column at the top. In the absorber, the sour gas comes into contact with the amine solution, where the acid gas components such as  $CO_2$  will be removed by the weak  $CO_2$ -amine bonding. Sweet gas then exits from the top of column while the rich amine solution is then fed into a

flash drum where hydrocarbons are vented out at the top vapour stream. The rich amine solution will then flow through the rich-lean-heat exchanger to be heated up and subsequently enters the stripper column. In the stripper column, the amine solution is regenerated by heating the solvent to strip off the acid gases at a low pressure, which then leaves the column from the top. The regenerated or lean amine solution exits from the bottom of column and transfers heat to the rich amine solution in the heat exchanger. Make-up amine is added to the lean amine stream before entering the absorber column (Behroozsarand and Zamaniya, 2011; Peters et al., 2011; Shakerian et al., 2015).



**Figure 4.2:** Simplified flow sheet of amine gas sweetening plant

## 4.4.1 Problem Formulation

### 4.4.1.1 Design Goal

The application of alkanolamines for the removal of acid gas from sour gas has long been practiced in the gas industry for almost 60 years. Some advantages of using alkanolamines include their high reactivity and low solvent cost. Among amines, aqueous monoethanolamine (MEA)



solution is the most commonly used solvent in gas sweetening process as it exhibits fast reaction rate, low solvent cost, reasonable thermal stability, and high absorbing capacity (Kumar et al., 2014). However, it also suffers from one major drawback, which is solvent loss due to its high volatility and degradation (Wang et al., 2015). Hence, this results in higher MEA make-up cost. The goal of this case study is to determine a solvent that will replace MEA as the absorbent which can help in minimising the usage of amine solution in the acid gas removal unit. Since MEA is known to be a nontoxic solvent, the generated solvent must not exhibit high safety and health hazard level. Hence, inherent safety and health indexes are applied as the evaluation tool to assess the safety and health performance of the generated molecule.

#### **4.4.1.2 Target Properties**

For this case study, the aim is to identify an amine-based solvent serving as CO<sub>2</sub>-absorbent involved in a chemical reaction that can help in reducing the amount of solvent loss in a gas sweetening process. The solvent must also possess a favourable safety and health characteristics to ensure that it does not bring much harm to the employees dealing with the process. The design objective of this work is to develop a molecule with low vapour pressure ( $VP$ ) to ensure that the solvent does not vaporise easily, which helps in minimising solvent loss. Besides, the solvent should also have low soil sorption coefficient ( $\log K_{oc}$ ) to prevent the accumulation of the escaping solvent in one place (Chemmanurgattuvalappil and Eden, 2013; Ng et al., 2014a). These two mentioned target properties are selected as the design objectives to be optimised. Next, the relevant property constraints are determined to ensure that the generated molecule exhibits the similar characteristics as the conventional amine solvents. The selected property constraints include heat of vaporisation ( $H_v$ ), liquid molar volume

( $V_m$ ), viscosity ( $\eta$ ), molecular weight ( $M_w$ ), boiling point ( $T_b$ ) and melting point ( $T_m$ ). All property ranges of the solvent at standard condition (298 K and 1 atm) are shown in Table 4.10. The lower and upper boundary values for  $H_v$ ,  $VP$ ,  $M_w$ ,  $\eta$ ,  $T_b$  and  $T_m$  are acquired from Kumar et al. (2014) while the lower and upper boundary values for  $V_m$  are taken from Chemmangattuvalappil and Eden (2013).

**Table 4.10:** Property targets for case study (gas sweetening)

Property	Lower bound	Upper bound
$H_v$ (kJ/mol)	50	528
$V_m$ (cm <sup>3</sup> /mol)	40	224
$M_w$ (g/mol)	60	250
$\eta$ (cP)	-	460
$T_b$ (°C)	111	350
$T_m$ (°C)	-65	25
$VP$ (mm Hg)	minimum	-
$\log K_{oc}$	minimum	-

## 4.4.2 Model Development

### 4.4.2.1 Property Prediction Model

In this stage, all properties that are considered as the target properties in Section 4.4.1.2 have to be calculated through property prediction methods. From Table 4.10,  $H_v$ ,  $V_m$ ,  $T_b$  and  $T_m$  can be predicted with GC models presented by Hukkerikar et al. (2012b), while  $\eta$  can be estimated by GC model developed by Conte et al. (2008) as shown in Table 4.11. Meanwhile, properties without available GC models can be estimated using correlations or empirical relationships.  $VP$  can be calculated from the

normal boiling point,  $T_b$  of the component using an empirical relationship as shown by the following equation (Sinha and Achenie, 2003):

$$\log VP = 5.58 - 2.7 \left( \frac{T_b}{T} \right)^{1.7} \quad (4.12)$$

where  $T$  is the temperature at standard condition (298 K).  $\log K_{oc}$  can be calculated through the correlation given in terms of octanol-water partition coefficient,  $\log K_{ow}$  which is shown by Equation (4.13) (Seth, et al., 1999).  $\log K_{ow}$  can be determined through GC model by Hukkerikar et al. (2012b) as shown in Table 4.11.

$$\log K_{oc} = 1.03 \log K_{ow} - 0.61 \quad (4.13)$$

**Table 4.11:** GC models for selected properties in the case study

Property $p$	$f(P)$ in Equation (2.3)	Universal constants
$H_v$ (kJ/mol)	$H_v - H_{v0}$	$H_{v0} = 9.6127$ kJ/mol
$V_m$ (cm <sup>3</sup> /mol)	$V_m - V_{m0}$	$V_{m0} = 16$ cm <sup>3</sup> /mol
$\eta$ (cP)	$\ln \eta$	-
$T_b$ (K)	$\exp(T_b/T_{b0})$	$T_{b0} = 244.5165$ K
$T_m$ (K)	$\exp(T_m/T_{m0})$	$T_{m0} = 143.5706$ K
$\log K_{ow}$	$\log K_{ow} - K_{ow0}$	$K_{ow0} = 0.4876$

### 4.4.3 Molecular Design

#### 4.4.3.1 Molecular Blocks

The molecular blocks selected in this case study are based on the conventional absorbents that are utilised in gas sweetening process. Some of the frequently used absorbents are monoethanolamine (MEA),

diethanolamine (DEA), triethanolamine (TEA), methyldiethanolamine (MDEA), diglycolamine (DGA), diisopropanolamine (DIPA) (Kumar et al., 2014), and diisopropylamine (Chemangattuvalappil and Eden, 2013). The selected molecular blocks include  $\text{CH}_3$ ,  $\text{CH}_2$ ,  $\text{CH}$ ,  $\text{OH}$ ,  $\text{CH}_2\text{O}$ ,  $\text{CH}_2\text{NH}_2$ ,  $\text{CH}_2\text{NH}$ ,  $\text{CHNH}$ ,  $\text{CH}_3\text{N}$ , and  $\text{CH}_2\text{N}$ . In this case study, only simple-structured molecules are considered, hence only first-order groups are utilised to construct the molecules. The binary variables of  $x$  and  $z$  in Equation (2.3) are set to zero. Since most conventionally used absorbents are acyclic compounds, the variable  $g$  in Equation (4.3) takes the value of one.

#### 4.4.3.2 Structural Constraints

In order to ensure that a structurally feasible molecule is formed without containing any free bonds, structural constraints shown by Equations (4.2) and (4.3) are applied.

#### 4.4.4 Optimisation Model

##### 4.4.4.1 Property Operator Targets

**Table 4.12:** Property operator targets for case study (gas sweetening)

Property $p$	$\Omega_p$	Lower bound	Upper bound
$H_v$	$H_v - H_{v0}$	40.3873	518.3873
$V_m$	$V_m - V_{m0}$	0.024	0.208
$M_w$	$M_w$	60	250
$\eta$	$\ln \eta$	-	6.1312
$T_b$	$\exp(T_b/T_{b0})$	4.8117	12.7879
$T_m$	$\exp(T_m/T_{m0})$	4.2623	7.9779

The target properties listed in Table 4.10 are then converted into their respective property operator,  $\Omega_p$  as shown in Table 4.12. The property operator is illustrated by the simple function  $f(P)$  for each target property  $p$ , which is exactly the left-hand side of Equation (2.3). For instance, the property operator of  $T_m$  is given by the function of  $\exp(T_m/T_{m0})$ , where  $T_m$  is given in K, while  $T_{m0}$  has a value of 143.5706 K (Hukkerikar et al., 2012b). By substituting the lower and upper bounds of  $T_m$  from Table 4.10 into  $\exp(T_m/T_{m0})$ , the calculated lower and upper property operator bounds of  $T_m$  are shown in Table 4.12.

#### 4.4.4.2 Fuzzy Optimisation

**Table 4.13:** Property operator targets for design objectives (gas sweetening)

Property $p$	$\Omega_p$	LB	UB	Design goal
$VP$	$\exp(T_b/T_{b0})$	5.2373	9.3078	Minimise $VP$ (or maximise $T_b$ )
$\log K_{oc}$	$\log K_{ow} - K_{ow0}$	-2.6284	3.7313	Minimise $\log K_{oc}$ (or $\log K_{ow}$ )

As stated in Section 4.4.1.2,  $\log K_{oc}$  and  $VP$  are the two target properties selected as the design objectives. The two objectives can be conflicting; hence fuzzy optimisation is performed to simultaneously optimise both objectives. To apply fuzzy optimisation, the linear membership functions from Equations (4.4) and (4.5) and the constraint from Equation (4.6) are used. In order to optimise these two design objectives, their property operator range must first be identified. As  $VP$  is estimated from  $T_b$ , the property operator for  $VP$  is expressed in terms of

that of  $T_b$ . While the property operator for  $\log K_{oc}$  is also represented in terms of that of  $\log K_{ow}$ . Both  $\log K_{ow}$  and  $VP$  are optimised separately (by maximising or minimising) to identify their respective upper and lower property operator bounds. The lower and upper bound values for both properties are shown in Table 4.13. The remaining properties in Table 4.12 are served as property constraints to be fulfilled. Both design objectives can then be represented by the linear membership functions as shown in Equations (4.14) and (4.15):

$$\frac{\Omega_{VP} - 5.2373}{9.3078 - 5.2373} = \lambda_{VP} \quad (4.14)$$

$$\frac{3.7313 - \Omega_{\log K_{oc}}}{3.7313 + 2.6284} = \lambda_{\log K_{oc}} \quad (4.15)$$

The overall objective is given by Equation (4.7), subjected to the following constraints:

$$\lambda_{VP} \geq \lambda \quad (4.16)$$

$$\lambda_{\log K_{oc}} \geq \lambda \quad (4.17)$$

The optimisation model becomes a mixed-integer linear programming (MILP). Integer cuts can be applied to enumerate alternative solutions.

#### 4.4.5 Performance Analysis

##### 4.4.5.1 Selection of Inherent Safety and Health Sub-indexes

The generated molecules from previous stage subsequently undergo the performance analysis stage to assess their safety and health

characteristics. The safety assessment is carried out using the sub-indexes of flammability ( $I_{FL}$ ) and explosiveness ( $I_{EX}$ ). Meanwhile, the selected health sub-indexes are comprised of viscosity ( $I_{\eta}$ ), material phase ( $I_{MS}$ ), volatility ( $I_V$ ), exposure limit ( $I_{EL}$ ), and acute health hazard ( $I_{AH}$ ). The allocation of the safety and health sub-index scores are provided in Section 4.3.5 (Tables 4.2, 4.3 and 4.5 to 5.9). The total penalty score that a molecule received is calculated using Equation (4.11).

#### 4.4.5.2 Property Prediction Models

As the application of safety and health sub-indexes involved several physicochemical properties, their corresponding property prediction models must be identified to assist with property estimation. The involved properties include  $F_p$ ,  $T_b$ ,  $UEL$ ,  $LEL$ ,  $\eta$ ,  $T_m$ ,  $PEL$ , and  $LD_{50}$  for acute oral toxicity. The GC models for  $T_b$ ,  $T_m$ , and  $\eta$  are provided in Table 4.11. As shown in Table 4.14, GC model developed Hukkerikar et al. (2012b) is employed to predict  $F_p$ , while GC models by Hukkerikar et al. (2012a) are applied to determine  $PEL$  and  $LD_{50}$  for acute oral toxicity.

**Table 4.14:** GC models for selected properties used in sub-indexes

Property $p$	$f(P)$ in Equation (2.3)	Universal constants
$F_p$ (K)	$F_p - F_{p0}$	$F_{p0} = 170.7058 \text{ K}$
$PEL$ (mol/m <sup>3</sup> )	$-\log PEL$	-
$LD_{50}$ (mol/kg)	$-\log LD_{50} - A_{LD50} - B_{LD50}M_W$	$A_{LD50} = 1.9372;$ $B_{LD50} = 0.0016$

The value of  $PEL$  calculated in Table 4.14 contains the unit of mol/m<sup>3</sup>. In order to convert the unit of  $PEL$  into ppm, the following equation is applied:

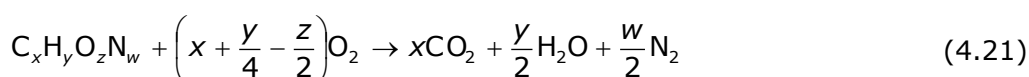
$$PEL' = V_{gas, std} \times PEL \times 1000 \quad (4.18)$$

where  $PEL'$  is expressed in terms of ppm while  $V_{gas, std}$  is the molar volume of gas or vapour at standard condition (298 K and 1 atm), which takes the value of 24.45 dm<sup>3</sup>/mol. Meanwhile, both lower explosion limit ( $LEL$ ) and upper explosion limit ( $UEL$ ) can be used interchangeably with lower flammability limit ( $LFL$ ) and upper flammability limit ( $UFL$ ) respectively. They can be calculated using the following correlations:

$$LFL = \frac{100\%}{1 + 9.0454C_o} \quad (4.19)$$

$$UFL = \frac{100\%}{1 + 1.1843C_o} \quad (4.20)$$

where  $C_o$  is the oxygen stoichiometric coefficient in a reaction (Ma et al., 2013). Consider a general compound  $C_xH_yO_zN_w$  undergoes a complete combustion in air:



$C_o$  can then be calculated by:

$$C_o = x + \frac{y}{4} - \frac{z}{2} \quad (4.22)$$

#### 4.4.6 Results and Discussions

Since this case study aims to replace MEA, the safety and health performance of the generated molecules must not perform inferior than



that of MEA. The properties of MEA are shown in Table 4.15. The values of  $T_b$ ,  $T_m$ ,  $F_p$ ,  $S$ , and  $LD_{50}$  are taken from MSDS provided by Sigma-Aldrich. The values of  $\log K_{ow}$  and  $\eta$  are extracted from the U.S. National Library of Medicine. Meanwhile,  $VP$  and  $PEL$  of MEA can be found in the Occupational Safety and Health Administration (OSHA) Occupational Chemical Database. Since  $\log K_{oc}$  of MEA is not available, it can be predicted using Equation (4.13), which returns a value of -1.96. All the safety and health sub-index scores for MEA are shown in Table 4.16.

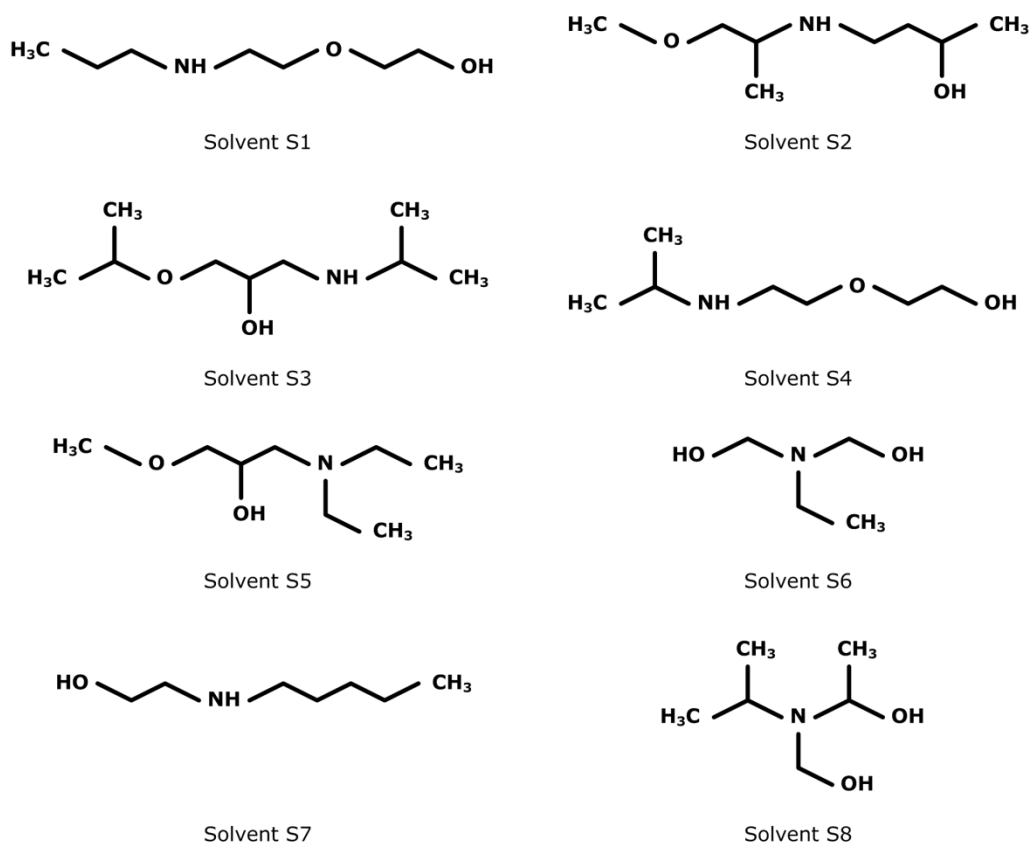
**Table 4.15:** Properties of MEA

Property	Property value	Property	Property Value
$\log K_{ow}$	-1.31	$T_m$ (°C)	10
$\log K_{oc}$	-1.96	$F_p$ (°C)	86
$VP$ (mm Hg)	0.4	$S$ (vol%)	14.5
$\eta$ (cP)	18.95	$PEL$ (ppm)	3
$T_b$ (°C)	170	$LD_{50}$ (mg/kg)	1720

**Table 4.16:** The sub-index scores of MEA

$I_{FL}$	$I_{EX}$	$I_{\eta}$	$I_{MS}$	$I_V$	$I_{EL}$	$I_{AH}$	$I_{SHI}$
2	1	3	2	0	3	1	12

The optimisation model is solved using LINGO 14.0 with a computational time of 0.4 seconds for the first generated solution. Eight solvents with the best  $\lambda$  values are generated, and their molecular structures are illustrated in Figure 4.3. Tables 4.17 and 4.18 demonstrate the properties of the eight generated solvents, which are all determined using property prediction models. From Table 4.17, all eight solvents have larger  $\log K_{oc}$  as compared to MEA. Note that the eight solvents are less



**Figure 4.3:** The best eight solvents with their molecular structures

**Table 4.17:** The eight generated solvents with their properties

Solvent	$\lambda$	$\log K_{oc} VP$			$H_v$	$V_m$	$M_w$
				(mm Hg)	(kJ/mol)	(cm <sup>3</sup> /mol)	(g/mol)
S1	0.597	-0.18	0.117		75.59	159.3	147.2
S2	0.587	-0.11	0.114		75.91	175.2	161.2
S3	0.548	0.14	0.071		78.44	191.4	175.3
S4	0.544	-0.37	0.185		73.37	159.0	147.2
S5	0.525	0.26	0.211		71.15	175.8	161.2
S6	0.518	-1.14	0.221		82.23	133.0	133.2
S7	0.513	-1.21	0.228		81.92	117.1	119.2
S8	0.506	0.42	0.232		71.76	153.0	131.2

**Table 4.18:** The eight generated solvents with their properties (continued)

Solvent	$\eta$	$T_b$	$T_m$	$F_p$	$S$	$PEL$	$LD_{50}$
	(cP)	(°C)	(°C)	(°C)	(vol%)(ppm)	(mg/kg)	
S1	16.32	227.0	24.8	115.4	6.544	0.666	662.5
S2	15.42	227.6	22.2	114.1	5.772	0.644	611.8
S3	16.66	236.9	23.6	119.2	5.163	0.582	579.9
S4	14.26	218.0	20.8	109.0	6.544	0.712	640.5
S5	7.51	215.4	-12.3	90.1	5.772	1.048	526.3
S6	36.66	214.4	21.1	107.6	7.553	2.443	771.0
S7	38.81	213.8	23.8	108.9	8.930	2.530	818.2
S8	10.93	213.4	16.2	94.0	6.265	4.772	744.6

**Table 4.19:** The eight generated solvents with their sub-index scores

Solvent	$I_{FL}$	$I_{EX}$	$I_{\eta}$	$I_{MS}$	$I_V$	$I_{EL}$	$I_{AH}$	$I_{SHI}$
S1	1	1	3	2	0	4	1	12
S2	1	1	3	2	0	4	1	12
S3	1	1	3	2	0	4	1	12
S4	1	1	3	2	0	4	1	12
S5	2	1	2	2	0	3	1	11
S6	1	1	3	2	0	3	1	11
S7	1	1	3	2	0	3	1	11
S8	1	1	3	2	0	3	1	11

volatile than MEA as they all have lower  $VP$  than MEA. Meanwhile, Table 4.19 shows all the sub-index values assigned to each molecule and their respective total index score. The best four solvents (S1 to S4) have an  $I_{SHI}$  value of 12 while the remaining four solvents have an  $I_{SHI}$  value of 11. Even though solvents S1 to S4 perform better in terms of target

functionality, their inherent safety and health performance are still inferior to that of solvents S5 to S8. This is mainly demonstrated by the highest  $I_{EL}$  penalty score received by solvents S1 to S4, which is deemed to be highly toxic when inhaled by the workers. In addition, solvents S5 to S8 also exhibit better safety and health performance compared to MEA, in which solvent S5 is less viscous than MEA, while solvents S6 to S8 are less flammable than MEA. Hence, solvents S5 to S8 can be considered as the chemical candidates for the gas sweetening process. The International Union of Pure and Applied Chemistry (IUPAC) name for solvents S5, S6, S7 and S8 are 1-(diethylamino)-3-methoxy-2-propanol, [n-ethyl(hydroxymethyl)amino]methanol, 2-(pentylamino)ethanol, and 1-[(hydroxymethyl)(isopropyl)amino]-1-ethanol respectively.

In this work, only the desired physicochemical properties of the molecule are considered as the design objectives in the early design stage. Safety and health aspects are only integrated in the performance analysis stage to evaluate the molecular performance. In order to ensure that the generated molecules exhibit both optimum targeted functionality and favourable safety and health performance, safety and health aspects should also be included as the design objectives along with the target physicochemical properties. This is addressed as the second scope of this research work and is presented in Chapter 5, which considers the simultaneous optimisation of molecular functional performance and its safety and health attributes.

In addition, the allocation of sub-index scores to the molecules is based on the property values, which are calculated via the property prediction models. These estimated property values are highly dependent on the accuracy of the property prediction models. Some estimated values

may have remarkable deviation from their respective actual experimental values. This may cause a lower or higher sub-index score to be assigned to certain molecules, since their property values may fall into another property sub-range resulting in inaccurate scores being assigned to the molecules. This issue has been identified as the third scope and is presented in Chapter 6 to enhance the safety and health sub-index scores under property prediction uncertainty.

One major limitation of the application of safety and health sub-indexes is the discontinuity of sub-index scores at the boundary values separating two adjacent sub-ranges. This causes the safety and health evaluation on the molecule to be less sensitive as property values that fall within the same interval are assigned a similar sub-index score. For instance, molecules with *PEL* value within 1 to 10 ppm are assigned a sub-index score of 3 while molecules with *PEL* value within 10 to 100 ppm are assigned a score of 2. Now consider that molecule *P* has a *PEL* of 3 ppm while molecule *Q* has a *PEL* of 8 ppm. Both molecules receive a similar sub-index score of 3, even though molecule *Q* has a higher *PEL* value, which indicates that it is inherently healthier than molecule *P*. Now consider another molecule *R* with *PEL* of 11 ppm, even though it has a *PEL* value difference of 3 ppm compared to molecule *Q*, it receives a score of 2. Meanwhile, molecule *P* and *Q* have a higher *PEL* value difference, but they both received the same score. Hence, one way to solve these issues is to modify the sub-index to ensure a continuous change in the scoring, which is addressed as the fourth scope and is presented in Chapter 7.

Nevertheless, the application of the existing safety and health indexes in CAMD in this work still provides a first insight on the consideration of safety and health aspects in the design stage. When these

inherently safer and healthier molecules are used in a process (eg. as solvent, catalyst, refrigerant, etc.), this helps to lower the overall hazard level present in the process.

## **4.5 Summary**

A novel methodology is proposed to assess the safety and health aspects of the optimal molecules generated by CAMD approach. The desired target properties of the molecular design problem are first identified in order to fulfil the customers' needs. These target properties can be estimated through property prediction models. The appropriate molecular building blocks are selected depending on the nature of the design problem. In order to ensure that a feasible molecule that does not contain any free bonds is produced, structural constraints are introduced into the problem. Optimisation is then carried out by optimising the design objectives subjected to property constraints to generate the optimal solution. Integer cuts are then introduced as constraints to generate alternative feasible solutions. The generated set of molecules then undergoes a performance analysis stage to evaluate their safety and health performance. Inherent safety and health indexes are applied as the assessment tool. Each molecule is allocated with several sub-index scores, and the summation of the scores is used to quantify and rank the inherent hazard level of each molecule. A case study on the solvent design for gas sweetening process by applying this method is illustrated. The objective is to design a solvent acting as absorbent to help in minimising the usage of amine solution in the acid gas removal unit. The design objectives of this problem are to minimise both the soil sorption coefficient and vapour pressure of the solvent subjected to the desired target property constraints. Eight solvents with optimised results are formulated and safety

and health evaluations are subsequently performed on these solvents. The two safety sub-indexes selected are flammability and explosiveness, while the five health sub-indexes selected include viscosity, material phase, volatility, exposure limit and acute health hazard. Sub-index scores are then assigned to each molecule and the total index score of each molecule is then calculated. From the results, the most optimal solvent does not possess the lowest index score. There exists a need to revise the current CAMD approach to include the aspects of safety and health in the early decision-making stage. This new approach can assist in designing molecules that are optimal in targeted property performance as well as safety and health performance.

# **CHAPTER 5**

## **A MOLECULAR DESIGN METHODOLOGY BY THE SIMULTANEOUS OPTIMISATION OF PERFORMANCE, SAFETY AND HEALTH ASPECTS**

### **5.1 Introduction**

In most CAMD problems, molecular physical and thermodynamic properties are often selected as the design criteria during design stage to ensure that the synthesised molecules fulfil the targeted functionalities. However, the incorporation of safety and health aspects into CAMD is not strongly emphasised as design criteria in many design problems. They are mainly introduced as property constraints so that molecules that do not fulfil the safety and health criteria are screened out. Instead of eliminating potential molecules, it is necessary to analyse the trade-off between the molecular functional performance and its safety and health characteristics. This chapter addresses the second scope of this research, in which both safety and health aspects are integrated as design criteria in the existing CAMD method to ensure that the synthesised molecule does not bring harm and health-related hazards to the consumers, and at the same time exhibits high functional performance. A novel chemical product design methodology has been developed to integrate both safety and health aspects, as well as the target physicochemical properties into a single-stage CAMD framework. The assessment of safety and health parameters



are based on the molecular properties that have significant impact on both aspects. Each property is introduced with a sub-index value depending on the degree of potential hazards. Disjunctive programming algorithm is employed to assist in converting the input molecular property values into their corresponding sub-index scores. Fuzzy optimisation is applied to optimise two principal design objectives in this work: molecular target properties and its safety and health performance. A case study on solvent design for gas sweetening process has been carried out to determine the optimal molecule with reasonably low safety and health hazards level, and at the same time, achieves the targeted property functionalities.

## 5.2 Problem Statement

Generally in most conventional CAMD problems, the design objectives to achieve are often the physicochemical properties of the molecule. These target properties are optimised to ensure that the generated molecule is able to attain the desired product functionality. The aspects of safety and health are usually imposed as design constraints so that the generation of hazardous molecules can be prevented. However, the 'enforcement' of such constraints may suppress the generation of solvents which excel dominantly in terms of product functionality, but are screened out due to not meeting the imposed constraints. For instance, given that a constraint is considered where  $F_p$  of the molecules is set to be 'greater than 50°C' to generate less flammable molecules. With this constraint, any molecules with  $F_p$  lower than 50°C are excluded from the final solutions. However, an excluded molecule with  $F_p$  of 49°C may have significantly superior functional performance as to that of a candidate molecule with  $F_p$  of 51°C, even though their  $F_p$  values are relatively close to one another. Instead of imposing constraints on the safety and health

factors, a trade-off should be conducted between the advantages in the performance of molecules and their safety and health attributes. In this chapter, inherent safety and occupational health of the molecule are served as the design criteria, along with the targeted properties of the molecule. The application of inherent safety and health indexes is used as the assessment tools to examine the molecular performance. There is one specific problem to be addressed, which is stated as follows:

1. In Chapter 4, the safety and health sub-indexes are employed after the optimal molecules are generated to assess the molecular safety and health performance. In this chapter, both targeted functional properties and the safety and health aspects of the molecules are simultaneously optimised. Thus, the allocation of sub-index scores to the molecules in this case has to be done concurrently when the optimisation model is searching for the optimal solution. Since the allocation of scores is based on the input property values, disjunctive programming algorithm is introduced to convert the properties into their corresponding scores.

## **5.3 Methodology**

The main scope of this work is to present a systematic framework that incorporates the aspects of safety and health as design criteria in a CAMD problem. This framework considers both product functionality and inherent safety and health properties simultaneously, in order to generate an optimal molecule with respect to both criteria.

### **5.3.1 Problem Formulation**

This stage serves the same purpose as the one presented in Section 4.3.1. The desired target properties which define the molecular targeted

functionalities and characteristics are identified. Each target property is bounded by a pre-specified range as given by Equation (4.1) to ensure the generated molecules can function and behave in the desired manner. Target properties to be optimised are selected as design objectives, while the remaining properties will act as property constraints.

### **5.3.2 Inherent Safety and Health Sub-index Selection**

In Chapter 4, both safety and health aspects are not considered as design objectives as they are only evaluated in the performance analysis stage once the optimal molecules are generated. Since the main goal of this chapter is to include both safety and health aspects as design criteria, both aspects are optimised simultaneously along with the targeted functional properties. Thus, the selection of safety and health sub-indexes are carried out before the molecules are generated. As explained in Section 4.3.5, the existing inherent safety and occupational health indexes are developed to assess and compare the hazard level of different possible chemical process routes to manufacture the same chemical compound. Each index is made up of numerous sub-indexes, which can be divided into chemical-related sub-indexes and process-related sub-indexes. The appropriate chemical-related sub-indexes have been selected in Section 4.3.5, which comprised of flammability ( $I_{FL}$ ) and explosiveness ( $I_{EX}$ ) for safety; and viscosity ( $I_{\eta}$ ), material phase ( $I_{MS}$ ), volatility ( $I_V$ ), exposure limit ( $I_{EL}$ ), and acute health hazard ( $I_{AH}$ ) for health. Each sub-index is evaluated by single or multiple properties, which are divided into several sub-ranges to represent different levels of potential hazard. A sub-range with higher degree of hazard is assigned a larger penalty score and vice versa. The sub-index scores for the seven chosen sub-indexes are shown in Tables 4.2, 4.3, and 4.5 to 4.9. The overall hazard level exhibited by a

molecule is quantified by summing up all the seven sub-index values as given by Equation (4.11).

### 5.3.3 Model Development

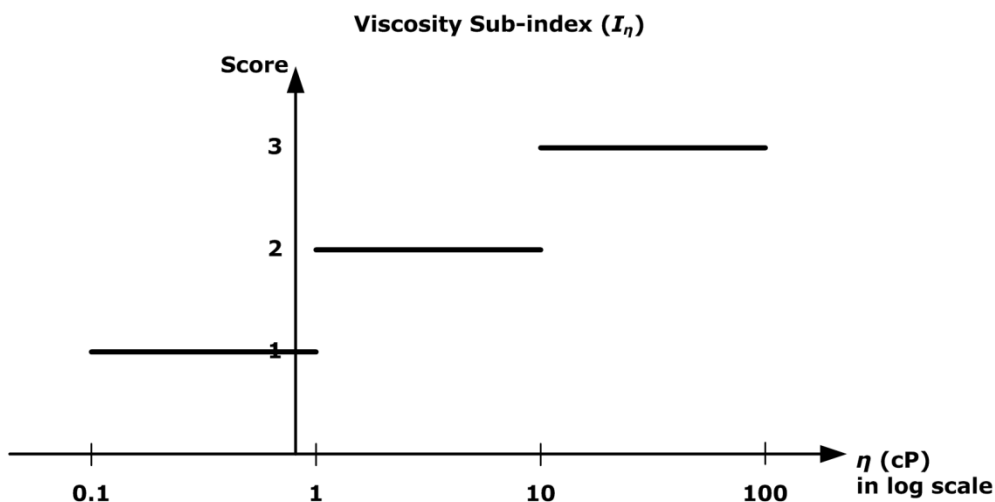
#### 5.3.3.1 Property Prediction Models

The molecular properties specified in Section 5.3.1 are the desired target properties for the generated molecules to fulfil the design goal. Meanwhile, the application of safety and health sub-indexes in Section 5.3.2 also involves the use of various safety and health-related properties. In this stage, all properties that are considered as target properties and in the safety and health sub-indexes have to be determined via property prediction models. As previously stated in Section 4.4.5.2, the GC models for  $T_b$ ,  $T_m$ , and  $\eta$  are given in Table 4.11, while the GC models for  $F_p$ ,  $PEL$ , and  $LD_{50}$  for acute oral toxicity are shown in Table 4.14. Both  $UEL$  and  $LEL$  can be calculated through the correlations provided by Equations (4.19) and (4.20).

#### 5.3.3.2 Disjunctive Programming for Allocation of Sub-index Scores

The allocation of penalty scores to the molecule is based on its input property values. All seven sub-indexes from Section 5.3.2 are comprised of several property sub-ranges, with each sub-range represented by a sub-index value. For instance, for the viscosity sub-index ( $I_\eta$ ) as shown in Figure 5.1, if the viscosity of a particular molecule falls in between 0.1 cP and 1 cP, a scoring of one is assigned to it. If the viscosity falls between 1 cP and 10 cP, then a scoring of two is allocated to it and so on. As the viscosity of 1 cP is a boundary separating the "0.1 to 1 cP" sub-range and "1 cP to 10 cP" sub-range, the scoring switches abruptly from one to two

as the viscosity moves across it. The presence of multiple sub-ranges in the sub-index has created a disjunction for the constraint. A modelling method known as disjunctive programming employs discontinuous functions to describe abrupt changes over a certain decision variable (El-Halwagi, 2012). Therefore, disjunctive programming can be adequately applied in this problem. As mentioned in Section 5.3.3.1, property prediction models are able to estimate the properties of interest for a molecule, while disjunctive programming inputs these properties into the safety and health sub-indexes in order to convert them into the corresponding sub-index scores. In this way, the optimisation model is able to simultaneously allocate scores to the potential generated molecules while searching for the optimal solution of a given design problem.



**Figure 5.1:** The graphical illustration of viscosity sub-index ( $I_\eta$ )

This section demonstrates the disjunctive algorithm that is employed in this work. Consider there is a particular sub-index score model which assigns a score of  $I_A$  when the corresponding property is lower than a boundary property value ( $p_{switch}$ ). Meanwhile, a score of  $I_B$  is allocated in

the event where the property value is equivalent to or above  $p_{switch}$ . This sub-index score model can be represented by Equation (5.1).

$$I_p = \begin{cases} I_A & p < p_{switch} \\ I_B & p \geq p_{switch} \end{cases} \quad (5.1)$$

Binary integer variables are introduced to model these functions. They are transformed to the following mixed-integer formulation using a binary integer variable ( $I$ ):

$$I_p = I_A * I + I_B * (1 - I) \quad (5.2)$$

subjected to the following condition:

$$I = \begin{cases} 0 & p \geq p_{switch} \\ 1 & p < p_{switch} \end{cases} \quad (5.3)$$

In order to ensure that the model assigns the correct value to  $I$  to satisfy condition (5.3), the following constraints have to be included:

$$(p_L - p_{switch}) * I < p - p_{switch} \leq (p_U - p_{switch}) * (1 - I) \quad I \in \{0,1\} \quad (5.4)$$

where  $p_L$  and  $p_U$  are the lower and upper bounds respective to any feasible  $p$  value. When  $p$  is smaller than  $p_{switch}$ , the term " $p - p_{switch}$ " becomes negative, forcing  $I$  to be 1 to satisfy both equalities in constraint (5.4). On the other hand, when  $p$  is greater than or equals to  $p_{switch}$ , the term " $p - p_{switch}$ " becomes positive, forcing  $I$  to be 0 to again satisfy both equalities in constraint (5.4). In this section, the application of disjunctive programming

on the index scoring is demonstrated for the viscosity sub-index. Consider the following criteria:

$$I_{\eta} = \begin{cases} 1 & 0.1 \text{ cp} \leq \eta < 1 \text{ cp} \\ 2 & 1 \text{ cp} \leq \eta < 10 \text{ cp} \\ 3 & 10 \text{ cp} \leq \eta \leq 100 \text{ cp} \end{cases} \quad (5.5)$$

The viscosity can be determined from the GC method in Table 4.11. Since the left hand side of its GC equation is given by the natural logarithm of  $\eta$ , it is crucial to reduce the complexity of the formulation model by keeping it as linear as possible. In order to eliminate these non-linear terms in the formulation, Equation (5.5) can be rewritten as shown below with the viscosity intervals represented in terms of  $\ln \eta$ :

$$I_{\eta} = \begin{cases} 1 & -2.3026 \leq \ln \eta < 0 \\ 2 & 0 \leq \ln \eta < 2.3026 \\ 3 & 2.3026 \leq \ln \eta \leq 4.6052 \end{cases} \quad (5.6)$$

The viscosity sub-index score  $I_{\eta}$  may be one, two or three depending on the viscosity of the molecule. Binary integer variables are used to model these functions. The  $I_{\eta}$  function can be transformed to the following mixed-integer formulation using two integer variables ( $I_{\eta 1}$  and  $I_{\eta 2}$ ):

$$I_{\eta} = I_{\eta 1} + I_{\eta 2} + 1 \quad (5.7)$$

subjected to the following conditions:

$$I_{\eta 1} = \begin{cases} 0 & \ln \eta < 0 \\ 1 & \ln \eta \geq 0 \end{cases} \quad (5.8)$$

$$I_{\eta 1} = \begin{cases} 0 & \ln \eta < 2.3026 \\ 1 & \ln \eta \geq 2.3026 \end{cases} \quad (5.9)$$

In order to model conditions (5.8) and (5.9) which assign the values of  $I_{\eta 1}$  and  $I_{\eta 2}$  to either 0 or 1 based on the viscosity of the molecule, the following constraints are considered:

$$-2.3026 * (1 - I_{\eta 1}) < \ln \eta \leq 4.6052 * I_{\eta 1} \quad (5.10)$$

$$-4.6052 * (1 - I_{\eta 2}) < \ln \eta - 2.3026 \leq 2.3026 * I_{\eta 2} \quad (5.11)$$

### 5.3.4 Molecular Design

In this stage, the first step is to identify the suitable first-order molecular groups serving as potential building blocks for molecular formation. The list of first-order molecular building blocks can be found in Hukkerikar et al. (2012a,b). In this chapter, second-order molecular groups are also considered as they are able to help differentiate distinct isomeric structures and improve the accuracy of the estimated properties. The possible second-order molecular groups are chosen based on the selection of first-order groups. Next, structural feasibility constraints are imposed to ensure that the generated molecules are feasible and do not contain any free bonds.

The mathematical algorithms presented by Zhang et al. (2015) in their CAMD problems are considered in this chapter. The following sets are first defined:

$$G_1 = \{i | i \text{ is a first-order group}\};$$



$ID = \{id | id \text{ is the ID number of each group}\};$

Then, several binary variables are introduced to describe the connectivity of molecular groups. Binary variable  $b_{i_1, id_1, i_2, id_2}$  signifies whether first-order group  $i_1$  with id  $id_1$  ( $i_1, id_1$ ) is connected to first-order group  $i_2$  with id  $id_2$  ( $i_2, id_2$ ), in which  $i_1, i_2 \in G_1$  and  $id_1, id_2 \in ID$ .

$$b_{i_1, id_1, i_2, id_2} = \begin{cases} 1 & \text{group}(i_1, id_1) \text{ is connected to group}(i_2, id_2) \\ 0 & \text{otherwise} \end{cases}$$

Meanwhile, another binary variable,  $z_{i_1, id_1}$  is assigned to represent the existence of group ( $i_1, id_1$ ) in the molecule.

$$z_{i_1, id_1} = \begin{cases} 1 & \text{group}(i_1, id_1) \text{ exists in the molecule} \\ 0 & \text{otherwise} \end{cases}$$

#### 5.3.4.1 Structural Constraints: First-Order Groups

In this section, the octet rule of structural feasibility is employed (Odele and Macchietto, 1993) to ensure that the synthesised molecules do not contain free bonds:

$$\sum_{i \in G_1} N_i (2 - v_i) = 2g \quad (5.12)$$

$$\sum_{i_1 \neq i_2; i_1, i_2 \in G_1} N_{i_1} \geq N_{i_2} (v_{i_2} - 2) + 2 \quad \forall i_2 \in G_1 \quad (5.13)$$

where  $v_i$  is the valence of group  $i$ ,  $N_i$  is the number of occurrence for first-order group  $i$ , and  $g$  is 1, 0, -1 or -2 for acyclic, monocyclic, bicyclic and tricyclic compounds respectively. Besides, the mathematical constraints

presented by Churi and Achenie (1996) are imposed to assure that only a single molecular structure is produced.

#### 5.3.4.2 Structural Constraints: Second-Order Groups

All second-order groups are made up by the connection of first-order groups using single bond (Zhang et al., 2015), and the full list is also provided in Hukkerikar et al. (2012a,b). The selection of second-order groups is decided by the first-order groups that are applied in a particular CAMD problem. For instance, the second-order group  $(\text{CH}_3)_2\text{CH}$  can only be present in a molecule when a single CH group is bonded to two  $\text{CH}_3$  groups. In order for  $(\text{CH}_3)_2\text{CH}$  to be chosen as one of the second-order groups, both first-order groups  $\text{CH}_3$  and CH must first be selected as the building blocks. Disjunctive programming can then be utilised to determine the number of second-order groups exists in a molecule. The disjunctive constraints below are applied for second-order group  $(\text{CH}_3)_2\text{CH}$ :

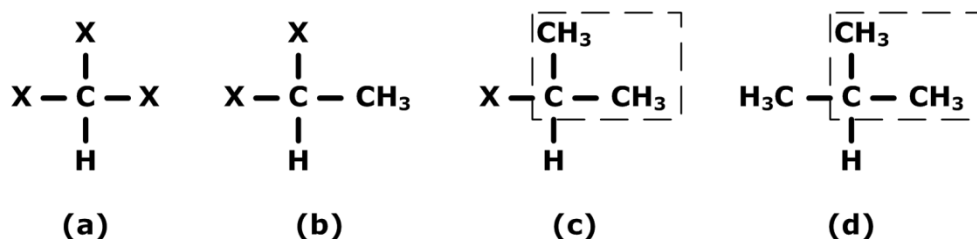
$$\left( L_{(\text{CH}_3)_2\text{CH}} - C_{(\text{CH}_3)_2\text{CH},id_1} \right) * \left( 1 - I_{(\text{CH}_3)_2\text{CH},id_1} \right) \leq \sum_{id_2 \in ID} b_{\text{CH},id_1,\text{CH}_3,id_2} - C_{(\text{CH}_3)_2\text{CH},id_1} \quad (5.14)$$

$$\sum_{id_2 \in ID} b_{\text{CH},id_1,\text{CH}_3,id_2} - C_{(\text{CH}_3)_2\text{CH},id_1} < \left( U_{(\text{CH}_3)_2\text{CH}} - C_{(\text{CH}_3)_2\text{CH},id_1} \right) * I_{(\text{CH}_3)_2\text{CH},id_1} \quad (5.15)$$

$$M_{(\text{CH}_3)_2\text{CH}} = \sum_{id_1 \in ID} I_{(\text{CH}_3)_2\text{CH},id_1} \quad (5.16)$$

where  $id_1, id_2 \in ID$ ,  $L_{(\text{CH}_3)_2\text{CH}}$  and  $U_{(\text{CH}_3)_2\text{CH}}$  represent the lower and upper limit for the number of  $\text{CH}_3$  groups bonded to CH group respectively,  $C_{(\text{CH}_3)_2\text{CH},id_1}$  represents a boundary value to denote whether  $(\text{CH}_3)_2\text{CH}$  is present in the molecule,  $I_{(\text{CH}_3)_2\text{CH},id_1}$  is a binary variable which denotes whether group CH with ID number  $id_1$  is connected to at least two  $\text{CH}_3$  groups, and  $M_{(\text{CH}_3)_2\text{CH}}$  represents the total number of  $(\text{CH}_3)_2\text{CH}$  group present in the molecule. Since a single CH group has a valence number of

three, hence the possible number of CH<sub>3</sub> groups that could be connected to CH group can either be zero, one, two or three. Hence,  $L_{(CH_3)_2CH}$  and  $U_{(CH_3)_2CH}$  are assigned the value of zero and three respectively.



**Figure 5.2:** The four possible scenarios where CH group is connected to (a) none, (b) one, (c) two or (d) three CH<sub>3</sub> group(s)

In order to form (CH<sub>3</sub>)<sub>2</sub>CH in a molecule, at least two CH<sub>3</sub> groups must be connected to a single CH group, as shown in Figure 5.2. When only one or none CH<sub>3</sub> is connected to the CH group, (CH<sub>3</sub>)<sub>2</sub>CH can never be formed. Hence, the boundary value,  $C_{(CH_3)_2CH,id1}$  should be assigned a value between one and two; and in this work, the value of 1.5 is allocated to  $C_{(CH_3)_2CH,id1}$ . In the event where the number of CH<sub>3</sub> groups connected to CH is greater than 1.5, (CH<sub>3</sub>)<sub>2</sub>CH group is present. Otherwise, the formation of (CH<sub>3</sub>)<sub>2</sub>CH group will not occur. Hence, Equations (5.14) and (5.15) can now be reduced to

$$-1.5 * (1 - I_{(CH_3)_2CH,id1}) \leq \sum_{id_2 \in ID} b_{CH,id1,CH_3,id2} - 1.5 \quad \forall id_1 \in ID \quad (5.17)$$

$$\sum_{id_2 \in ID} b_{CH,id1,CH_3,id2} - 1.5 < 1.5 * I_{(CH_3)_2CH,id1} \quad \forall id_1 \in ID \quad (5.18)$$

For each CH group in the molecule, if the number of CH<sub>3</sub> groups connected to that particular CH group is greater than 1.5, the value of  $I_{(CH_3)_2CH,id1}$  is forced to become one to fulfil constraints (5.17) and (5.18).

Otherwise, the value of  $I_{(\text{CH}_3)_2\text{CH},id1}$  becomes zero. All values of  $I_{(\text{CH}_3)_2\text{CH},id1}$  for each CH group are then be summed up as shown in Equation (5.16) to determine the number of occurrence of  $(\text{CH}_3)_2\text{CH}$  in the molecule.

### 5.3.5 Optimisation Model

In this chapter, the two main criteria to be optimised are the functionality performance of the molecule and its safety and health aspects performance. In order to ensure an inherently safer and healthier molecule is generated, the total penalty score of the safety and health sub-indexes,  $I_{SHI}$  must be minimised. However, the molecule which performs better in terms of functionality may not necessarily exhibit a low penalty score. Hence, a decision making has to be made on the trade-off between the target performance of the molecule and its inherent safety and health level. The design criteria here are conflicting in nature and fuzzy optimisation algorithm can thus be applied to ensure that both desirable product functionality and the safety and health criteria can be attained. Similarly to Section 4.3.4, a degree of satisfaction for target property,  $\lambda_p$  has to be introduced to each target property selected as design objective, while a degree of satisfaction for inherent safety and health,  $\lambda_I$  is applied to  $I_{SHI}$ .  $\lambda_p$  can be represented by the linear membership functions as given by Equations (4.4) and (4.5). As for  $\lambda_I$ , it can be expressed as a linear membership function bounded by lower and upper bounds of  $I_{SHI}$  as shown in Equation (5.19), in which  $I_{SHI}$  is to be minimised. The lower bound ( $I_{SHI}^L$ ) and upper bound ( $I_{SHI}^U$ ) of  $I_{SHI}$  are determined by the lowest and highest  $I_{SHI}$  values respectively subjected to the target property constraints and structural constraints.

$$\lambda_{I_{SHI}} = \frac{I_{SHI}^U - I_{SHI}}{I_{SHI}^U - I_{SHI}^L} \quad (5.19)$$

All  $\lambda_p$  and  $\lambda_I$  can be simultaneously optimised by the max-min aggregation approach presented by Zimmermann (1978), which maximises the least satisfied degree of satisfaction. To ensure Pareto optimality, the objective function of this optimisation model is given by Equation (5.20), which is expanded from Equation (4.7), subjected to constraints (4.8) and (5.21). Once the first optimal molecule is generated, integer cuts are then added to generate alternative optimal molecules.

$$\max \lambda + \frac{1}{M} \left( \lambda_{I_{SHI}} + \sum_{p \in P} \lambda_p \right) \quad (5.20)$$

$$\lambda_{I_{SHI}} \geq \lambda \quad (5.21)$$

## 5.4 Case Study: Solvent Design for Gas Sweetening Process

### 5.4.1 Problem Formulation

The proposed methodology in Section 5.3 is carried out on the same case study as given in Section 4.4. The design goal of this CAMD problem is to identify an amine-based solvent serving as CO<sub>2</sub>-absorbent involved in a chemical reaction that can help in reducing the amount of solvent loss in a gas sweetening process. The proposed methodology also focuses on the aspects of safety and occupational health, in which both the solvent performance and the safety and health attributes are simultaneously considered as design criteria when developing the solvents. Three properties are chosen as design objectives to be optimised, which include  $VP$ ,  $\log K_{oc}$  and  $I_{SHI}$ . The former two properties represent the desired

solvent characteristics, while the latter property measures the inherent safety and health level of the solvent. The desired solvent candidate must have low  $VP$  so that it does not vaporise easily, low  $\log K_{oc}$  to prevent the accumulation of the escaping solvent in one place, and low  $I_{SHI}$  to ensure that it is inherently safer and healthier. Meanwhile,  $H_v$ ,  $V_m$ ,  $\eta$ ,  $M_w$ ,  $T_b$ , and  $T_m$  are selected as property constraints and their respective target range are shown in Table 4.10.

### 5.4.2 Inherent Safety and Health Sub-index Selection

One of the design objectives is to minimise  $I_{SHI}$ , which is the sum of the seven selected sub-indexes that include  $I_{FL}$ ,  $I_{EX}$ ,  $I_{\eta}$ ,  $I_{MS}$ ,  $I_V$ ,  $I_{EL}$ , and  $I_{AH}$ .

### 5.4.3 Model Development

In this stage, the property prediction models for all involved properties in this case study are identified. All GC models are summarised in Tables 4.11 and 4.14, while the correlations are provided by Equations (4.12), (4.13), (4.19), and (4.20). Disjunctive programming as demonstrated in Section 5.3.3.2 is then employed to allocate penalty scores for all the seven sub-indexes.

### 5.4.4 Molecular Design

In order to generate the molecules, the same first-order molecular blocks as listed in Section 4.4.3.1 are chosen. Meanwhile, the appropriate second-order molecular groups are selected based on the first-order groups, which include  $(CH_3)_2CH$ ,  $CH(CH_3)CH(CH_3)$ ,  $CHOH$ ,  $CH_a(OH)CH_b(OH)$ , and  $CH_a(OH)CH_b(NH_\gamma)$ , where  $a$ ,  $b$  and  $\gamma$  can either be zero, one or two. The binary variable of  $x$  in Equation (2.3) is now set to

one for the inclusion of second-order groups. The structural constraints discussed in Sections 5.3.4.1 and 5.3.4.2 are applied.

### 5.4.5 Optimisation Model

All property constraints listed in Section 5.4.1 are transformed into their respective property operator,  $\Omega_p$  as provided in Table 4.12. Unlike the case study in Section 4.4, this case study has three design objectives comprising of  $VP$ ,  $\log K_{oc}$  and  $I_{SHI}$ . In order for these three properties to be optimised, their property operator range must first be determined. Each of these three properties is optimised one at a time (either by maximising or minimising) to identify their respective upper and lower property operator bounds. The lower and upper bound values (LB and UB respectively) for these three design objectives are summarised in Table 5.1.

**Table 5.1:** Property operators targets for design objectives (gas sweetening)

Property $p$	$\Omega_p$	LB	UB	Design goal
$VP$	$\exp(T_b/T_{b0})$	5.2373	9.6504	Minimise $VP$ (or maximise $T_b$ )
$\log K_{oc}$	$\log K_{ow} - K_{ow}$	-2.6284	3.5933	Minimise $\log K_{oc}$ (or $\log K_{ow}$ )
$I_{SHI}$	$I_{SHI}$	10	13	Minimise $I_{SHI}$

All objectives are then expressed by linear membership functions as shown in Equations (5.22) to (5.24):

$$\frac{\Omega_{VP} - 5.2373}{9.6504 - 5.2373} = \lambda_{VP} \quad (5.22)$$

$$\frac{3.5933 - \Omega_{\log K_{oc}}}{3.5933 + 2.6284} = \lambda_{\log K_{oc}} \quad (5.23)$$

$$\frac{13 - \Omega_{I_{SHI}}}{13 - 10} = \lambda_{I_{SHI}} \quad (5.24)$$

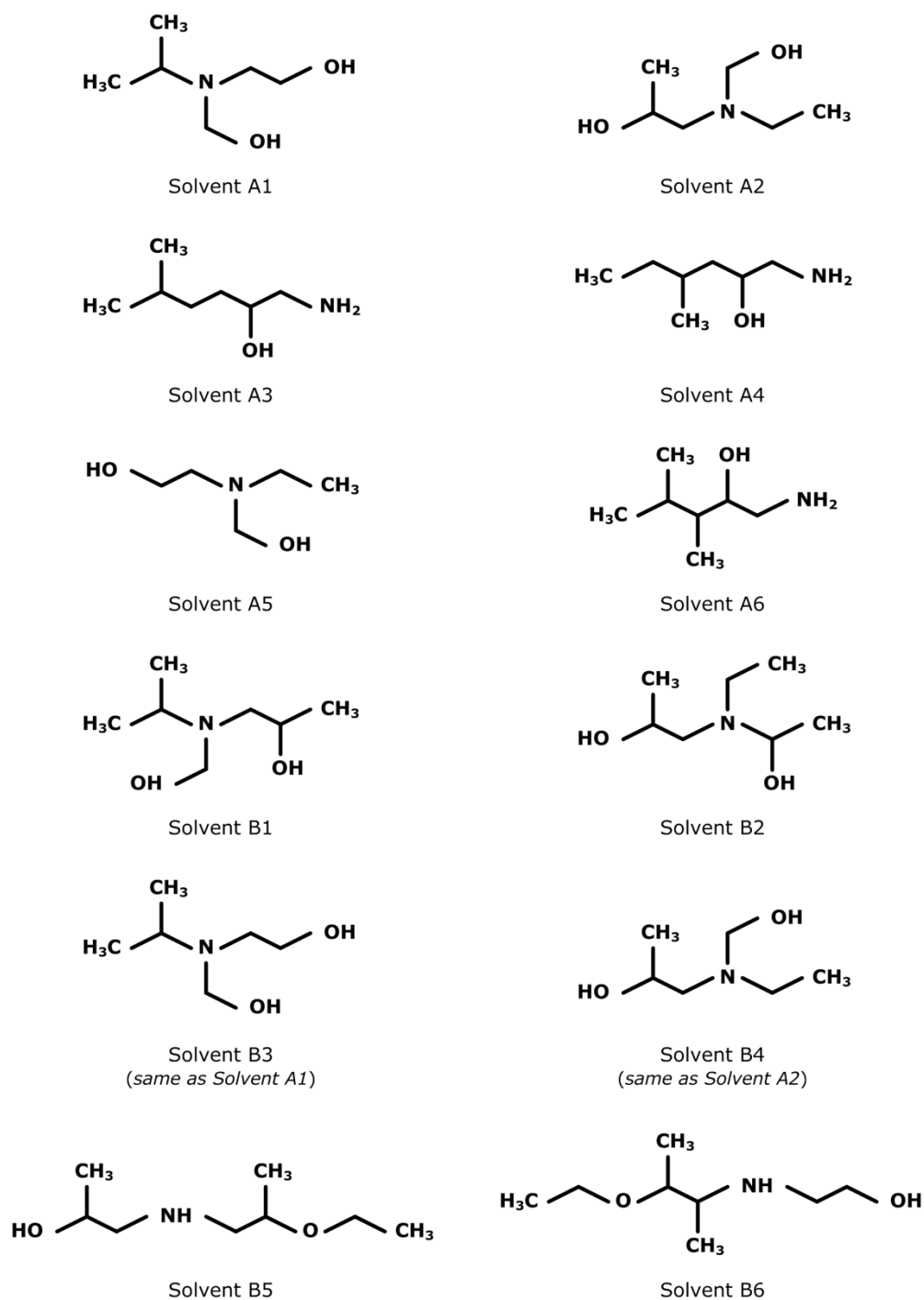
The overall objective function is given by Equation (5.20), subjected to constraints (4.16), (4.17), and (5.21). The optimisation model is a mixed-integer nonlinear programming (MINLP) due to the formulation of  $I_{AH}$ . The property used to assess this sub-index is  $LD_{50}$  (acute oral toxicity), in which a logarithm variable is present in the GC model. In this case study, two different design problem scenarios are considered in the results. In Scenario A, the design objectives are to optimise  $VP$ ,  $\log K_{oc}$  and  $I_{SHI}$ . In scenario B, the safety and health criteria are not considered as the design objectives. Thus, only  $VP$  and  $\log K_{oc}$  are optimised, while  $I_{SHI}$  of each molecule is only calculated after the optimised molecules are generated. Scenario B works the same as the research scope presented in Chapter 4, but the results may be different as second-order molecular groups are applied in this chapter. Both scenarios are conducted to examine the effect of integration of the safety and health sub-indexes has on the generated molecules.

#### 5.4.6 Results and Discussions

The optimisation model is solved using LINGO 14.0 with a computational time of 98 minutes for the first generated solution. Six solvents with the six highest  $\lambda$  values for each scenario are generated, and their molecular structures are illustrated in Figure 5.3. The six molecules generated in Scenario A are labelled as solvent A1 to A6; while the six molecules in Scenario B are named as solvent B1 to B6. Tables 5.2 and 5.3 show the properties of the six generated solvents in Scenario A, while



Tables 5.5 and 5.6 show the properties of the six generated solvents in Scenario B. The individual sub-index scores of the generated solvents in Scenarios A and B are summarised in Tables 5.4 and 5.7 respectively. It is noted that the solvents that are ranked third and fourth (Solvents B3 and B4) in Scenario B are ranked first and second (Solvents A1 and A2) in Scenario A. This is because among the six solvents generated in Scenario B, solvents B3 and B4 exhibit better safety and health performance as displayed by their lower  $I_{SHI}$  scores in Table 5.7. Hence, the integration of safety and health aspects as design criteria (as demonstrated by Scenario A) helps to prioritise molecules with improved safety and health characteristics. By comparing Scenarios A and B, the methodology in Scenario B is easier to compute in an optimisation model, since the model only optimises  $VP$  and  $\log K_{oc}$ . As  $I_{SHI}$  can be manually calculated after the optimal molecules are generated, disjunctive programming algorithm shown in Section 5.3.3.2 is not necessarily required. In Scenario B, six iterations with the addition of integer cuts have been conducted to form six different solvents. However, when these six solvents are ranked according to their  $I_{SHI}$ , the best solution among them is not guaranteed to be the global optimal solution. More iteration is thus needed to increase confidence level for the global optimality of the solution, which is rather time-consuming and indefinite. Therefore, the optimisation approach presented in Scenario A is preferable as it simultaneously optimise  $VP$ ,  $\log K_{oc}$  and  $I_{SHI}$ . In this way, the solution generated from the first iteration is guaranteed to be the global optimum solution. Hence, the six solvents generated in Scenario A are the top six global solutions with respect to property functionality and safety and health aspects. The discussion from this point onwards focuses only on Solvents A1 to A6, unless it is stated otherwise.



**Figure 5.3:** The generated solvents with their molecular structures

**Table 5.2:** The generated solvents with their properties in Scenario A

Solvent $\lambda$	$\log K_{oc}$ $VP$			$H_v$	$V_m$	$M_w$
	(mmHg)			(kJ/mol)	(cm <sup>3</sup> /mol)	(g/mol)
A1	0.614	-0.93	0.083	84.22	133.6	133.2
A2	0.612	-1.05	0.084	84.45	132.7	133.2
A3	0.559	-0.34	0.122	74.01	147.2	131.2
A4	0.558	-0.46	0.124	74.24	146.3	131.2
A5	0.545	-1.31	0.136	81.92	116.5	119.2
A6	0.533	-0.51	0.147	72.65	144.8	131.2

**Table 5.3:** The generated solvents with their properties in Scenario A  
(continued)

Solvent $\eta$	$T_b$	$T_m$	$F_p$	$S$	$PEL$	$LD_{50}$
(cP)	(°C)	(°C)	(°C)	(vol%)(ppm)	(mg/kg)	
A1	78.52	233.7	22.5	119.2	7.553	1.057 898
A2	77.42	233.5	21.7	119.0	7.553	1.006 836
A3	25.24	226.2	23.7	100.7	6.265	2.437 1022
A4	24.88	225.9	23.0	100.5	6.265	2.320 952
A5	71.63	224.1	20.3	113.9	8.930	1.113 858
A6	33.15	222.5	22.7	97.8	6.265	2.578 887

**Table 5.4:** The generated solvents with their sub-index scores in Scenario A

Solvent	$I_{FL}$	$I_{EX}$	$I_{\eta}$	$I_{MS}$	$I_V$	$I_{EL}$	$I_{AH}$	$I_{SHI}$
A1	1	1	3	2	0	3	1	11
A2	1	1	3	2	0	3	1	11
A3	1	1	3	2	0	3	1	11
A4	1	1	3	2	0	3	1	11
A5	1	1	3	2	0	3	1	11
A6	1	1	3	2	0	3	1	11

**Table 5.5:** The generated solvents with their properties in Scenario B

Solvent	$\lambda$	$\log K_{oc}$ $VP$		$H_v$	$V_m$	$M_w$
			(mm Hg)	(kJ/mol)	(cm <sup>3</sup> /mol)	(g/mol)
B1	0.666	-0.67	0.052	86.76	149.8	147.2
B2	0.653	-0.84	0.063	88.25	148.5	147.2
B3	0.614	-0.93	0.083	84.22	133.6	133.2
B4	0.612	-1.05	0.084	84.45	132.7	133.2
B5	0.594	-0.21	0.069	75.91	174.6	161.2
B6	0.591	-0.19	0.053	76.77	172.5	161.2

**Table 5.6:** The generated solvents with their properties in Scenario B (continued)

Solvent	$\eta$ (cP)	$T_b$ (°C)	$T_m$ (°C)	$F_p$ (°C)	$S$ (vol%)	$PEL$ (ppm)	$LD_{50}$ (mg/kg)
B1	84.87	242.7	23.9	124.3	6.544	0.955	865
B2	67.71	239.0	23.0	125.2	6.544	0.730	766
B3	78.52	233.7	22.5	119.2	7.553	1.057	898
B4	77.42	233.5	21.7	119.0	7.553	1.006	836
B5	28.45	237.3	18.6	119.2	5.772	0.283	641
B6	42.76	242.3	21.7	122.7	5.772	0.280	576

**Table 5.7:** The generated solvents with their sub-index scores in Scenario B

Solvent	$I_{FL}$	$I_{EX}$	$I_{\eta}$	$I_{MS}$	$I_V$	$I_{EL}$	$I_{AH}$	$I_{SHI}$
B1	1	1	3	2	0	4	1	12
B2	1	1	3	2	0	4	1	12
B3	1	1	3	2	0	3	1	11
B4	1	1	3	2	0	3	1	11
B5	1	1	3	2	0	4	1	12
B6	1	1	3	2	0	4	1	12

From the generated results as shown in Tables 5.2 and 5.3, solvent A1 displays the lowest  $VP$  value while solvent A5 has the lowest  $\log K_{oc}$  value. According to Table 4.15, MEA has  $VP$  and  $\log K_{oc}$  values of 0.4 mm Hg and -1.96 respectively. All solvents have a relatively lower  $VP$  compared to MEA, but their  $\log K_{oc}$  values are still larger than that of MEA. In comparison of their inherent safety and health performance as provided in Table 5.4, all six solvents received a similar  $I_{SHI}$  score of 11. The similarity

in  $I_{SHI}$  scoring is due to the fact that all safety and occupational health sub-indexes are represented by multiple property sub-ranges, with each sub-range allocated with a discrete value. The safety and health-related property values of the solvents fall in the same corresponding sub-ranges, thus resulting in the similarity of scores for all individual sub-indexes. Even though all six solvents have high penalty scores for the sub-indexes of viscosity and exposure limit, they offer low flammability, low explosiveness, very low volatility and low acute health hazard.

It was reported that inhalation is the main source of chemical exposure occupationally since the respiratory system is the most common route for chemical contaminants in the form of gas, vapour and fume to enter the human body (Hassim and Hurme, 2010a). Highly volatile chemicals are more likely to be inhaled by the workers in the event of an accidental leakage. According to the volatility sub-index  $I_v$ , Table 4.7 shows that molecules with normal boiling point ( $T_b$ ) above 150°C are considered to have very low volatility. From Table 5.2, the six solvents have  $T_b$  values of over 220°C, which is significantly higher than the threshold of 150°C. Even though the generated solvents possess high exposure limit impact (high  $I_{EL}$  score), they do not vaporise easily to bring about any airborne disease to the employees. As a result, the application of these optimal molecules in process plants can help in minimising the adverse safety and occupational health impacts resulted from the hazards associated to chemicals.

Meanwhile, the accuracy and reliability of the property prediction models play a crucial in this work. As the sub-index scores are allocated based on the estimated property values, the accuracy of the assigned sub index scores is significantly affected by the accuracy of the prediction

models. Hence, the comparison study between the estimated properties with their corresponding experimental property values should be conducted. However, the experimental property data for all generated solvents may not be readily available. This is due to the fact that these solvents are not commercially available in the market. One way to predict the experimental property values of a molecule with novel molecular structure is to search for the data from another commercially available molecule that has the same molecular formula and functional groups. However, certain property data for the latter molecule may still not be available. For instance, both solvents A1 and A2 have the similar molecular formula of  $C_6H_{15}NO_2$ . Another commercially available chemical with the same molecular formula is diisopropanolamine, and its property data can be taken from PubChem. The only known properties are its  $F_p$  (127°C),  $T_b$  (248°C),  $UEL$  (5.4 vol%),  $LEL$  (1.1 vol%), and  $LD_{50}$  (4765 mg/kg). As its  $PEL$  and  $\eta$  values at standard condition are not available, it hinders the calculation for the final  $I_{SHI}$  of diisopropanolamine. Therefore, future work can be proposed to carry out verification on the molecular properties and performance of the generated solvents through experimental work. Besides, the uncertainty resulted from the property prediction models can also be analysed to study its effect on the accuracy of the sub index scores allocated to the molecules, which is the main research scope for the next chapter.

## 5.5 Summary

In this work, a single-stage CAMD framework has been developed to design molecule with low safety and health risks level that also achieves a set of desired target properties. The existing safety and health indexes are adapted into the CAMD problem to evaluate the safety and health

attributes of the molecules. Disjunctive programming algorithm has been integrated into the framework for converting the input property values into their respective sub-index scores. The calculation of total index score of a molecule enables users to quantify and compare its inherent hazard level. Fuzzy optimisation is then employed to simultaneously optimise multiple design objectives: product functionality and safety and health performance. A case study on the solvent design for a gas sweetening process is carried out to develop amine-based solvents that simultaneously achieve high functionality and favourable safety and health characteristics. The results show that the proposed methodology is able to generate molecules that achieve the desired product functionality and also possess high safety and health performance. Since the allocated index scores are highly dependent on the molecular properties, the accuracy of the property prediction models has high impact on the accuracy of the index scores. The following chapter considers the enhancement for the accuracy of the index scoring to better reflect the inherent hazard level of a molecule under property prediction uncertainty.



# **CHAPTER 6**

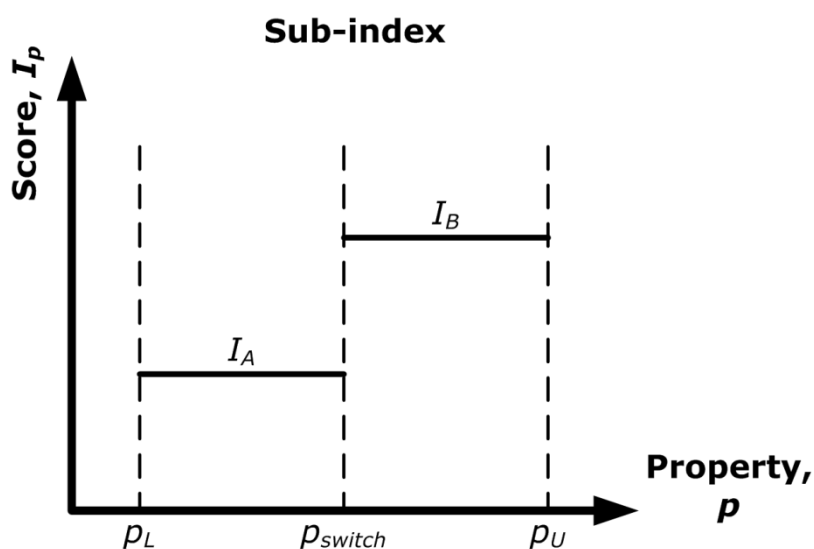
## **MANAGING UNCERTAINTY ON THE INTEGRATION OF SAFETY AND HEALTH INDEXES IN COMPUTER-AIDED MOLECULAR DESIGN**

### **6.1 Introduction**

In Chapter 5, a chemical product design methodology has been presented to integrate both safety and health aspects into the CAMD framework. The measurement of safety and health performance are based on the molecular properties that have impacts on both of these aspects. In all sub-indexes, the properties involved are divided into few sub-ranges, where each sub-range is assigned with a sub-index value or penalty score reflecting the degree of potential hazards. This approach ensures that a molecule that satisfies the targeted properties, and at the same time meets the safety and health criteria, is synthesised. In CAMD, all involved molecular properties are estimated through property prediction methods. Property prediction models offer the advantage of quick estimation without the need of conducting empirical test to identify the property values of interest. Thus, the reliability and effectiveness of these prediction models in estimating the properties significantly affect the accuracy of the allocated score to the molecules. However, uncertainty resulted from the utilisation of property prediction models may adversely affect the accuracy of scores assigned to the molecules. As the allocation of scores serve as the safety and health indicators, uncertainties are managed on the safety

and health sub-indexes to enhance the allocated scores for improved measurement of inherent hazard level demonstrated by the molecules. A case study on solvent design for carotenoid extraction from palm pressed fibre (PPF) has been carried out under property prediction uncertainty to determine the optimal molecule with reasonably low safety and health hazards level and optimum functionality.

## 6.2 Problem Statement



**Figure 6.1:** The allocation of sub-index scores based on property sub-ranges

Currently, the commonly used approach to assign sub-index score is by dividing the property into several sub-ranges, where each sub-range is represented by a discrete value. This approach is rather user-friendly as one can rapidly determine the individual sub-index score without involving any mathematical formula. However, the use of multiple property sub-ranges has a limitation, where the score switches abruptly at the property boundary separating two adjacent sub-ranges. This limitation becomes notable as uncertainty from property prediction models are considered in

the sub-indexes. In order to demonstrate this issue, the same example as given in Section 5.3.3.2 is applied, where there is a particular sub-index score model which assigns a score of  $I_A$  when the property is lower than the property boundary ( $p_{switch}$ ), while a score of  $I_B$  is allocated in the event where the property is equivalent to or above  $p_{switch}$ . This example of sub-index model is illustrated as shown Figure 6.1,  $p$  represents the property value,  $I_p$  denotes the sub-index score, while  $p_L$  and  $p_U$  are the feasible property lower and upper bounds respectively.

The molecular physicochemical properties can be swiftly estimated using property prediction methods without the need of carrying out experimental work. However, the discrepancies between the actual experimental values and estimated values for most property prediction models are around 10% or higher. Since the scores assigned to the molecule are based on the estimated property values, the accuracy of the allocated scores is thus dependent on the accuracy of the prediction models. The deviation of the estimated property value from the actual value may result in an inaccurate score being assigned to the molecule. This issue is especially more significant when the predicted value is near the property boundary,  $p_{switch}$  as shown in Figure 6.1. This is the point where the sub-index score switches from  $I_A$  to  $I_B$  as  $p$  moves from the lower score property sub-range (below  $p_{switch}$ ) to the adjacent higher score sub-range (above  $p_{switch}$ ). For example, given that a molecule  $M$  has an estimated  $p$  that is slightly higher than  $p_{switch}$ , and its given  $I_p$  score according to Figure 6.1 would be  $I_B$ . As  $p$  is estimated by prediction model, it is possible that its actual property value is lower or higher than the estimated value within an acceptable range. Thus, it is possible that the actual value itself is slightly lower than  $p_{switch}$ . In the case where the actual  $p$  is below  $p_{switch}$ , then its  $I_p$  score should now be  $I_A$ . Thus, it is observed

that the property boundary region is deemed to be highly uncertain and sensitive to the assignment of sub-index scores. Uncertainty resulted from the accuracy of property prediction model can cause the score to be shifted to a different value. In this chapter, property prediction uncertainty is managed on the safety and health sub-indexes, where the main concern is the uncertainty in allocating scores at the property boundary region. The outcome of this chapter is to ensure that the scores can be adjusted and enhanced to better represent the inherent hazard level of a molecule under prediction uncertainty. There are several specific problems to be addressed, which are stated as follows:

1. As the allocation of scores at all property boundaries are deemed to be highly uncertain, an uncertain range must be determined for each property boundary to manage the uncertainty. One way is to identify and apply the statistical performance indicators such as standard deviation and average absolute error provided by the property prediction models to determine the uncertain range.
2. As the sub-index scores in the uncertain range are modified to account for property prediction uncertainty, the scores are then expressed in terms of different functions. Thus, disjunctive programming is adopted to model the score functions for the conversion of property values into their respective scores.

## **6.3 Methodology**

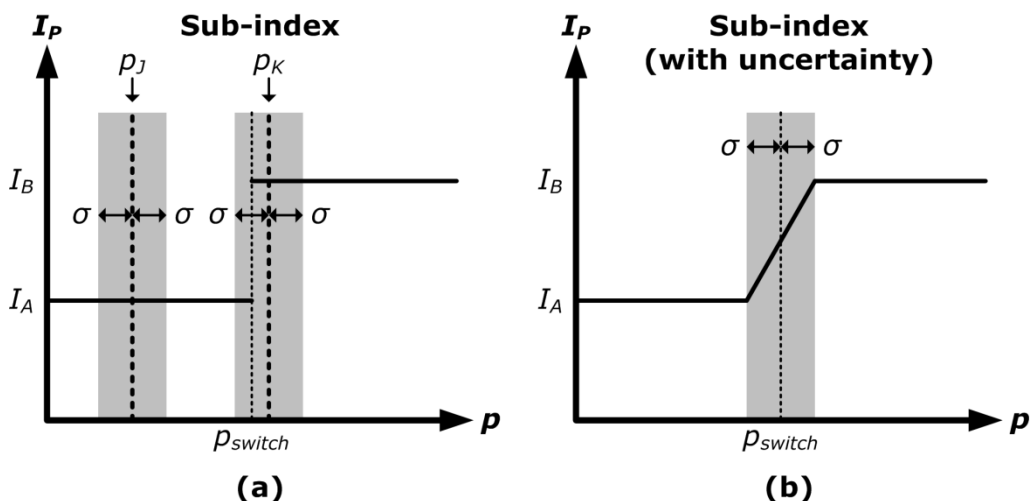
The methodology proposed in this chapter is an extension for the work done in Chapter 5. As shown in Figure 3.3, the additional stage considered in this chapter is the uncertainty management in the sub-indexes. Meanwhile, the disjunctive programming algorithm used in this chapter differs from that of the previous chapter (as presented in Section

5.3.3.2), where score functions are applied to account for property prediction uncertainty. Hence, the uncertainty management stage is illustrated in Section 6.3.1, while the modified disjunctive programming is demonstrated in Section 6.3.2.

### 6.3.1 Managing Uncertainty in Sub-indexes

Prior to this phase, the CAMD steps involved in this methodology are as follow: problem formulation, selection of safety and health sub-indexes, and identification of property prediction models. The full descriptions for these three steps are presented in Sections 5.3.1, 5.3.2 and 5.3.3.1 respectively. In this stage, the issue of uncertainty is addressed by illustrating it with the application of the simple sub-index model example as given in Figure 6.1. In this example, given that there are two molecules,  $J$  and  $K$ , with estimated property values of  $p_J$  and  $p_K$  respectively as shown in Figure 6.2(a). Their corresponding  $I_p$  scores would be  $I_A$  and  $I_B$  respectively. For all property prediction models, their capability to estimate the properties of chemical components are usually expressed in terms of statistical performance indicators. Consider that the prediction model for this particular property has a standard deviation of  $\sigma$ , then the actual experimental property value for molecule  $J$  can fall within the range of  $p_J \pm \sigma$ , which is represented by the grey region on the left in Figure 6.2(a). In this range, the  $I_p$  score for  $p_J$  is always fixed at  $I_A$ . Thus, uncertainty from the prediction model does not affect the sub-index score assigned to molecule  $J$ . Meanwhile, the actual property value for molecule  $K$  is within the range of  $p_K \pm \sigma$ , as represented by the grey region on the right in Figure 6.2(a). However, the  $I_p$  score in this range is uncertain as it can either be  $I_A$  or  $I_B$ . This is due to the fact that the property boundary,  $p_{switch}$ , happens to fall in the uncertain range for molecule  $K$ . As  $p_{switch}$  is the point that separates two adjacent sub-ranges with different  $I_p$  scores, any

molecule with estimated  $p$  value that is near to  $p_{switch}$  will encounter the same uncertainty issue with its score allocation. Hence, property prediction uncertainty must be managed mainly at the property boundary region.

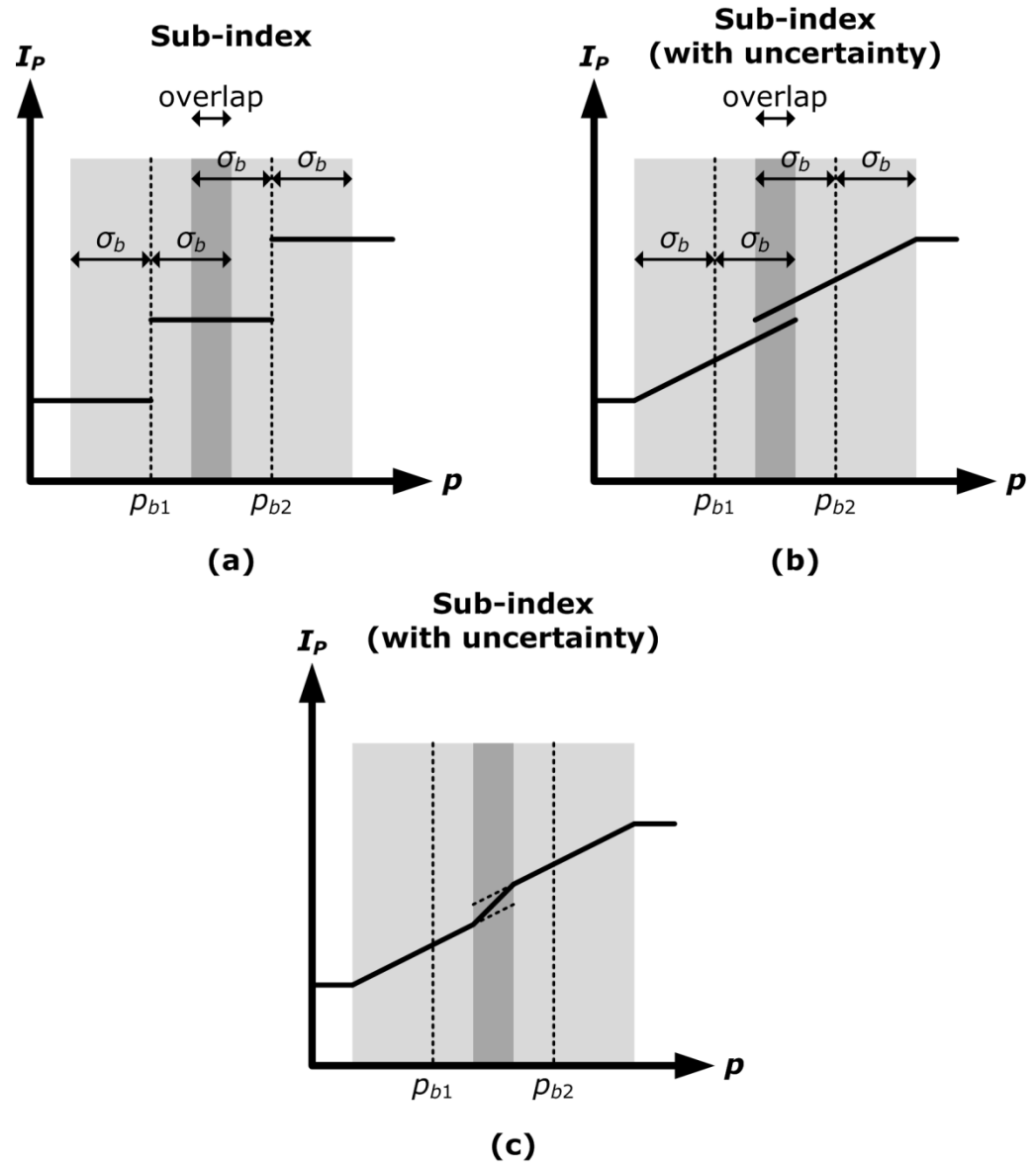


**Figure 6.2:** (a) Initial form of sub-index; (b) Revised sub-index with the incorporation of uncertainty

In order to address uncertainty in this sub-index, the standard deviation of the prediction model for this property,  $\sigma$  is added to or subtracted from  $p_{switch}$  to create the uncertain range  $[p_{switch} \pm \sigma]$ , the grey zone in Figure 6.2(b)]. At  $p_{switch} - \sigma$ , the initial score is  $I_A$ ; whereas at  $p_{switch} + \sigma$ , the initial score would be  $I_B$ . In this uncertain range, the  $I_p$  score transits linearly from  $I_A$  (at  $p_{switch} - \sigma$ ) to  $I_B$  (at  $p_{switch} + \sigma$ ). This transition is represented by the linear slope drawn on the uncertain range.

Another sub-index with different scorings is depicted in Figure 6.3. Given that the standard deviation of property prediction for the property is  $\sigma_b$ . As there are two property boundaries ( $p_{b1}$  and  $p_{b2}$ ) present in this particular sub-index, the uncertainty ranges at the two boundaries are determined as shown in Figure 6.3(a). However, there exists an overlapping of uncertain range for both boundaries (in the range of " $p_{b2} -$

$\sigma_b$  to " $p_{b1} + \sigma_b$ ") as denoted by the darker grey region in Figure 6.3(a). By using the similar linear transition slope in Figure 6.2(b), two linear slopes are drawn on the uncertain range for both boundaries.



**Figure 6.3:** (a) Initial form of sub-index; (b) Sub-index with the incorporation of uncertainty; (c) Revised sub-index with composite curve

As shown in Figure 6.3(b), the overlapping region now contains two different transition slopes contributed by each property boundary. The transition slope with the lower scores is contributed by the lower boundary,

$p_{b1}$ , while the other slope is by  $p_{b2}$ . However, it is meaningless if the  $p$  value at this overlapping region is given two different sub-index scores. A single property value at any feasible  $p$  region can only receive a single sub-index score to indicate its inherent hazard level. In order to address this issue, a composite curve is applied on this overlapping region. Both transition slopes in this region increase linearly from the lower boundary to the upper boundary of the dark grey region. At the lower boundary of the dark grey region ( $p_{b2} - \sigma_b$ ), the lowest  $I_p$  score is provided by the lower transition slope. Meanwhile, the highest  $I_p$  score at the upper boundary of the dark grey region ( $p_{b1} + \sigma_b$ ) is provided by the higher transition slope. A new composite curve is added by linearly connecting the lowest  $I_B$  score (at the lower boundary) to the highest  $I_B$  score (at the upper boundary) in the dark grey region. This resulting composite curve and the modified sub-index slopes are illustrated in Figure 6.3(c). The two methods proposed in Figures 6.2 and 6.3 are then applied to manage the property prediction uncertainty present in the safety and health sub-indexes.

### **6.3.2 Allocation of Sub-index Scores with Disjunctive Programming**

Once all the sub-indexes are modified by incorporating uncertainty, the property values of a molecule have to be translated into their corresponding scores. For instance, the revised sub-index as shown in Figure 6.2(b) shows that when  $p$  is below  $p_{switch} - \sigma$ , a score of  $I_A$  is allocated to the molecule. When  $p$  falls between  $p_{switch} \pm \sigma$ , the score given to the molecule is in the range of  $I_A$  to  $I_B$ , which can be represented by a linear function. When  $p$  is above  $p_{switch} + \sigma$ , the given score would be  $I_B$ . The score model for this particular sub-index in Figure 6.2(b) can be expressed by Equation (6.1):



$$I_p = \begin{cases} I_A & p \leq p_{switch} - \sigma \\ \frac{p - (p_{switch} - \sigma)}{2\sigma} (I_B - I_A) + I_A & p_{switch} - \sigma \leq p \leq p_{switch} + \sigma \\ I_B & p \geq p_{switch} + \sigma \end{cases} \quad (6.1)$$

As shown in Equation (6.1), the three  $p$  intervals have their scores represented by different functions. These intervals have resulted in a disjunction for the constraint. Disjunctive programming algorithm as presented in Section 5.3.3.2 is applied to model the different  $I_p$  score functions as given by Equation (6.1). Binary integer variables are used to model these functions. The  $I_p$  function can be transformed to the following mixed-integer formulation using two binary integer variables ( $b_1$  and  $b_2$ ):

$$I_p = b_1 I_A + (b_2 - b_1) \left[ \frac{p - (p_{switch} - \sigma)}{2\sigma} (I_B - I_A) + I_A \right] + (1 - b_2) I_B \quad (6.2)$$

subjected to the following conditions:

$$b_1 = \begin{cases} 0 & p \geq p_{switch} - \sigma \\ 1 & p < p_{switch} - \sigma \end{cases} \quad (6.3)$$

$$b_2 = \begin{cases} 0 & p \geq p_{switch} + \sigma \\ 1 & p < p_{switch} + \sigma \end{cases} \quad (6.4)$$

When  $p$  is smaller than  $p_{switch} - \sigma$ , both  $b_1$  and  $b_2$  take the value of one. According to Equation (6.2), its  $I_p$  value would then be equivalent to  $I_A$ . When  $p$  is larger than  $p_{switch} + \sigma$ , both  $b_1$  and  $b_2$  take the value of zero, its  $I_p$  value would then be  $I_B$ . When  $p$  is between  $p_{switch} \pm \sigma$ ,  $b_1$  and  $b_2$  would take the values of zero and one respectively, resulting in the calculated  $I_p$  value in Equation (6.2) to be expressed by the linear function representing the linear transition slope. Thus, the two binary variables in Equation (6.2) function as a switch to 'activate' the only  $I_p$  function of interest. In order to

ensure that the model assigns the correct values to  $b_1$  and  $b_2$  to satisfy conditions (6.3) and (6.4), the following constraints have to be included:

$$[p_L - (p_{switch} - \sigma)] * b_1 < p - (p_{switch} - \sigma) \leq [p_U - (p_{switch} - \sigma)] * (1 - b_1) \quad (6.5)$$

$$[p_L - (p_{switch} + \sigma)] * b_2 < p - (p_{switch} + \sigma) \leq [p_U - (p_{switch} + \sigma)] * (1 - b_2) \quad (6.6)$$

where  $p_L$  and  $p_U$  are the lower and upper bounds respective to any feasible  $p$  value. When  $p$  is smaller than  $p_{switch} - \sigma$ , the term " $p - (p_{switch} - \sigma)$ " in constraint (6.5) and " $p - (p_{switch} + \sigma)$ " in constraint (6.6) become negative, forcing  $b_1$  and  $b_2$  to be 1 to satisfy both equalities in constraints (6.5) and (6.6) respectively. On the other hand, when  $p$  is greater than or equals to  $p_{switch} + \sigma$ , the term " $p - (p_{switch} - \sigma)$ " in constraint (6.5) and " $p - (p_{switch} + \sigma)$ " in constraint (6.6) become positive, forcing  $b_1$  and  $b_2$  to be 0 to again satisfy both in constraints (6.5) and (6.6) respectively. When  $p$  is between  $p_{switch} \pm \sigma$ ,  $b_1$  and  $b_2$  are forced to become 0 and 1 respectively.

Once this step has been formulated in the optimisation model, the following steps are molecular design and optimisation model formulation. As the incorporation of uncertainty has resulted in the use of many non-linear equations, it is preferable to simplify the formulation of optimisation model. Thus, only first-order molecular groups are considered in this work. The full description of the phases of molecular design and optimisation model formulation are presented in Sections 4.3.3 and 5.3.5 respectively.

## 6.4 Case Study: Solvent Design for Extraction of Carotenoids

According to Oil World, palm oil is the world most important vegetable oil, as it accounts for 38.7% (62.6 million tons) of the vegetable

oil production in 2015 (European Palm Oil Alliance, 2016). The oil palm tree exhibits the highest yielding oil crop per unit area of cultivated land, in which the oil yield per hectare of plantation is ten times higher than other leading oilseed crops. Palm oil is utilised predominantly in edible food industry, while also applied in non-food industry such as soap and oleochemical manufacturing (Mba et al., 2015). After the extraction of crude palm oil from fresh fruit bunches (FFB), the remaining by-product is known as the palm pressed fibber (PPF). Nowadays, PPF is usually burned as fuel to supply energy for palm oil mills (Neoh et al., 2011), transported to the plantation along with empty fruit bunch (EFB) for field mulching (Lau et al., 2008), or used as animal feed (Dal Prá et al., 2016).

Choo et al. (1996) discovered that the residual oil found in PPF contains a significant amount of carotenoids (4000-6000 ppm), vitamin E (2400-3500 ppm), and sterols (4500-8500 ppm). The four major constituents of identified carotenoids include  $\beta$ -carotene,  $\alpha$ -carotene, lycopene, and phytoene. The quantity of carotenoids present in the residual oil in PPF is about six times higher than found in crude palm oil (França and Meireles, 1997). Carotenoids are mainly used in medical, cosmetic and biotechnological purposes, while also serve as natural colouring agents in food processing industry (Yara-Varón et al., 2016b).

One of the established approaches to recover carotenoids from PPF is the solvent extraction method (Neoh et al., 2011). It offers high performance in the recovery of residual oil and does not require frequent maintenance (Anderson, 2011). Among all solvents, *n*-hexane is the top choice as it offers high residual oil recovery and low polarity, and it can be easily separated from the products via evaporation. It has a moderate and convenient boiling point (68.5°C), which is appealing for the extraction

process since it is not highly volatile to cause any solvent loss (Yara-Varón et al., 2016a). Besides, it only requires low heat consumption for solvent recovery as it also offers low sensible heat (Anderson, 2011). However, hexane is deemed to be highly flammable, toxic to aquatic life, and may cause fatality if swallowed or enters airways.

#### **6.4.1 Problem Formulation**

In this case study, the design goal is to substitute hexane with an alternative solvent to recover carotenoids from PPF. The replacement solvent must exhibit lower flammability than hexane and must not cause severe destruction to aquatic life. In addition, it must not cause safety and health concerns to the on-site workers operating solvent extraction and recovery process. As hexane possesses an appropriate boiling point for solvent extraction and offers high solubility of carotenoids, the generated solvent must also demonstrate the similar attributes. The four design objective properties chosen in this case study are as follow:

1. the boiling point of hexane (68.5°C) is applied as the target benchmark for the boiling point ( $T_b$ ) of solvent. In this case study, the boiling point difference ( $T_{b,diff}$ ) between hexane and the developed solvent is minimised.
2. the heat of vaporisation ( $H_v$ ) of solvent is minimised for lower energy consumption during solvent recovery process.
3. the solubility of carotenoids in the solvent is maximised to achieve higher carotenoid extraction, which can be determined using the Hansen Solubility Parameters (HSP) (Hansen 2007).
4. the total index score ( $I_{SHI}$ ) is minimised for an inherently safer and healthier solvent.

As the main drawbacks of using hexane are its highly flammability and toxicity, properties such as flash point ( $F_p$ ), octanol-water partition coefficient ( $\log K_{ow}$ ), and acute toxicity (96-h  $LC_{50}$  to fathead minnow) are chosen as property constraints to ensure that the generated solvents do not demonstrate such undesirable characteristics. The upper and lower boundary values (UB and LB respectively) of the property constraints are given in Table 6.1.

**Table 6.1:** Property constraints for case study (carotenoid extraction)

Property	Standard deviation/error of GC method	LB	UB
$F_p$ ( $^{\circ}\text{C}$ )	12.10	-1.2	-
$\log K_{ow}$	0.64	-	2.86
$\log LC_{50}$ (96-h)	0.37	1.37	-

For  $F_p$ , the lower bound is decided based on the flash point of hexane ( $-23.3^{\circ}\text{C}$ ). In order to generate a molecule with lower flammability than hexane, its  $F_p$  must be set higher than that of hexane. As  $F_p$  is estimated using GC method developed by Hukkerikar et al. (2012b), the standard deviation of the  $F_p$  GC model should be taken into account while deciding on the lower bound. The standard deviation value for this model as given in Table 6.1 is  $12.1^{\circ}\text{C}$ . An additional  $F_p$  margin of  $10^{\circ}\text{C}$  is also incorporated to ensure the  $F_p$  of the generated molecule does not come close to that of hexane. Using  $F_p$  of hexane as the reference value, the lower bound  $F_p$  value for the proposed solvent is calculated by adding the temperature margin and standard deviation value to  $-23.3^{\circ}\text{C}$ , which returns a value of  $-1.2^{\circ}\text{C}$  as shown in Table 6.1. As for  $\log K_{ow}$  and  $LC_{50}$ , both properties are essential in identifying the environmental fate of the chemicals. The boundary value for  $\log K_{ow}$  is determined using the hazard

ranking criteria by Cordella et al. (2009). In this hazard ranking criteria,  $\log K_{ow}$  is divided into three sub-ranges with different degree of hazards. The sub-range with the least hazard level is chosen, which states that  $\log K_{ow}$  value of a chemical must be lower than 3.5. The same  $\log K_{ow}$  constraint is also applied by Patel et al. (2010) for the design of solvent. With this constraint, hexane will not emerge as one of the potential solvent candidates as it possesses a  $\log K_{ow}$  value of 3.94. This property can also be predicted using GC method by Hukkerikar et al. (2012b), with a standard deviation of 0.64. As for acute toxicity  $LC_{50}$ , its boundary value is determined based on the United Nations' Globally Harmonised System of Classification and Labelling of Chemicals (GHS). The lower bound for  $LC_{50}$  (96-h) in this case study is set at 10 mg/l. Hexane has a  $LC_{50}$  (96-h to fathead minnow) value of 2.5 mg/l, thus it does not fulfil the lower bound constraint and will not be generated. This property can be estimated using GC method proposed by Martin and Young (2001), with a root-mean-square error (RMSE) of 0.37 (expressed in terms of  $\log LC_{50}$ ). By considering the standard deviation or error of the prediction models, the revised boundary values for  $\log K_{ow}$  and  $\log LC_{50}$  are presented in Table 6.1.  $K_{ow}$  is unitless while  $LC_{50}$  has a unit of mg/l.

#### **6.4.2 Selection of Inherent Safety and Health Sub-indexes**

Similar to the case studies demonstrated in Sections 4.4 and 5.4, the seven sub-indexes applied to assess the molecular performance are  $I_{FL}$ ,  $I_{EX}$ ,  $I_{\eta}$ ,  $I_{MS}$ ,  $I_V$ ,  $I_{EL}$ , and  $I_{AH}$ . The sub-index scores are not the same as the ones used in the previous case studies, as uncertainty from property prediction is taken into account in the sub-indexes. The revised sub-index

scores are given in Section 6.4.3.2, which addresses the management of uncertainty on properties evaluated in sub-indexes.

### 6.4.3 Model Development

#### 6.4.3.1 Property Prediction Models

**Table 6.2:** GC models for selected properties in the case study (carotenoid extraction)

Property $p$	$f(P)$ in Equation (2.3)	Universal constants
$\delta_d$	$\delta_d$	-
$\delta_p$	$\delta_p$	-
$\delta_h$	$\delta_h$	-
$LC_{50}$ (mg/l)	$-\log LC_{50}$	-
$UEL$ or $UFL$ (vol%)	$\log (UFL/UFL_{const})$	$UFL_{const} = 129.9552$ vol%
$LEL$ or $LFL$ (vol%)	$\log (LFL/LFL_{const})$	$LFL_{const} = 4.5315$ vol%

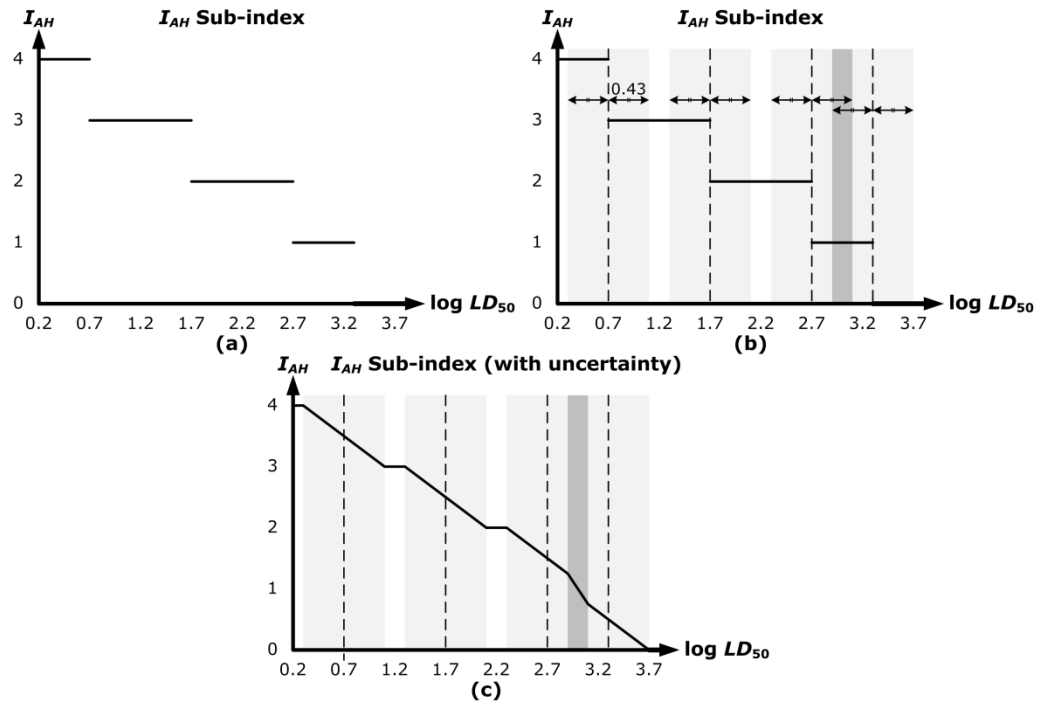
In Section 6.4.1, the involved target properties in this case study are  $T_b$ ,  $H_v$ , HSP parameters ( $\delta_d$ ,  $\delta_p$  and  $\delta_h$ ),  $F_p$ ,  $\log K_{ow}$ , and  $LC_{50}$  (96-h to fathead minnow). GC models to determine  $T_b$ ,  $H_v$ ,  $F_p$ , and  $\log K_{ow}$  have been provided in Tables 4.11 and 4.14. The three HSP parameters and  $LC_{50}$  are estimated by GC models developed by Hukkerikar et al. (2012b) and Martin and Young (2001) respectively. Meanwhile, both  $UEL$  and  $LEL$  (from  $I_{EL}$  sub-index) in this case study are predicted using GC model presented by Frutiger et al. (2016a). The prediction equations for HSP parameters,  $\log LC_{50}$ ,  $UEL$ , and  $LEL$  are given in Table 6.2.

#### 6.4.3.2 Managing Uncertainty in Sub-indexes

In this section, the management of uncertainty in acute health hazard sub-index ( $I_{AH}$ ) is demonstrated. Figure 6.4(a) illustrates the initial form of  $I_{AH}$  sub-index as provided by NFPA hazard health rating. This sub-index is evaluated using  $\log LD_{50}$  for acute oral toxicity, which can be estimated using the GC method by Hukkerikar et al. (2012a). First, the standard deviation for this GC model is identified as 0.43. Then, all property boundaries present in this sub-index must be determined, which are  $\log LD_{50}$  at 0.7, 1.7, 2.7, and 3.3. In Figure 6.4(b), all the uncertain ranges are highlighted as shown by the grey regions. Note that there is an overlapping region between  $\log LD_{50}$  at 2.87 and 3.13 as shown by the dark grey region. All linear transition slopes are drawn in the uncertain regions while composite curve is added in the overlapping region. Figure 6.4(c) shows the revised  $I_{AH}$  sub-index curves after taking into account property prediction uncertainty.

For other health sub-indexes excluding  $I_{MS}$ , the properties involved are  $T_b$ ,  $\eta$ , and  $PEL$ , where the standard deviations of their GC models are 7.9 K, 3.44 cP, and 0.78 respectively (Hukkerikar et al., 2012b; Conte et al., 2008; Hukkerikar et al., 2012a). These standard deviations are then used to determine the uncertain range of each property boundary. For each boundary, the upper bound of its uncertain range is the addition of the corresponding standard deviation to the boundary value. The lower bound can be determined by subtracting the corresponding standard deviation from the boundary value. A similar approach to apply transition slope as illustrated in Figure 6.4 is introduced to each uncertain range.





**Figure 6.4:** (a) Initial form of  $I_{AH}$  sub-index; (b)  $I_{AH}$  sub-index with the integration of uncertainty; (c) Revised  $I_{AH}$  sub-index curve

**Table 6.3:** Viscosity ( $I_\eta$ ) sub-index (revised form from PRHI) (Hassim and Edwards, 2006)

Parameter	Score Information	Penalty Score
Viscosity, $I_\eta$	$0.1 \text{ cp} \leq \eta < 1 \text{ cp}$	$(\eta + 1.7)/1.8$
	$1 \text{ cp} \leq \eta < 4.44 \text{ cp}$	$(\eta + 9.32)/6.88$
	$4.44 \text{ cp} \leq \eta < 6.56 \text{ cp}$	2
	$6.56 \text{ cp} \leq \eta < 13.44 \text{ cp}$	$(\eta + 7.2)/6.88$
	$13.44 \text{ cp} \leq \eta \leq 100 \text{ cp}$	3

**Table 6.4:** Volatility ( $I_V$ ) sub-index (revised form from IOHI) (Hassim and Hurme, 2010)

Parameter	Score Information	Penalty Score
Volatility, $I_V$	<i>Liquid and gas</i>	
	$T_b > 157.9^\circ\text{C}$	0
	$157.9^\circ\text{C} \geq T_b > 142.1^\circ\text{C}$	$(157.9 - T_b)/15.8$
	$142.1^\circ\text{C} \geq T_b > 57.9^\circ\text{C}$	1
	$57.9^\circ\text{C} \geq T_b > 42.1^\circ\text{C}$	$(73.7 - T_b)/15.8$
	$42.1^\circ\text{C} \geq T_b > 7.9^\circ\text{C}$	2
	$7.9^\circ\text{C} \geq T_b > -7.9^\circ\text{C}$	$(39.5 - T_b)/15.8$
	$T_b \leq -7.9^\circ\text{C}$	3

**Table 6.5:** Exposure limit ( $I_{EL}$ ) sub-index (revised form from IOHI) (Hassim and Hurme, 2010)

Parameter	Score Information	Penalty Score
Exposure limit, $I_{EL}$	<i>Vapour (ppm)</i>	
Let $P' = \log PEL$	$P' > 3.78$	0
	$2.78 < P' \leq 3.78$	$0.641(3.78 - P')$
	$2.22 < P' \leq 2.78$	$1.2821(2.78 - P') + 0.641$
	$1.78 < P' \leq 2.22$	$0.641(2.22 - P') + 1.359$
	$1.22 < P' \leq 1.78$	$1.2821(1.78 - P') + 1.641$
	$0.78 < P' \leq 1.22$	$0.641(1.22 - P') + 2.359$
	$0.22 < P' \leq 0.78$	$1.2821(0.78 - P') + 2.641$
	$-0.78 < P' \leq 0.22$	$0.641(0.22 - P') + 3.359$
	$P' \leq -0.78$	4

**Table 6.6:** Acute health hazard ( $I_{AH}$ ) sub-index (revised form from National Fire Protection Association, 2007)

Parameter	Score Information	Penalty Score
Acute health hazard, $I_{AH}$	<i>Oral rat LD<sub>50</sub> (mg/kg)</i>	
Let $D' = \log LD_{50}$	$D' > 3.731$	0
	$3.129 < D' \leq 3.731$	$1.1628(3.731 - D')$
	$2.871 < D' \leq 3.129$	$2.3256(3.129 - D') + 0.7$
	$2.269 < D' \leq 2.871$	$1.1628(2.871 - D') + 1.3$
	$2.129 < D' \leq 2.269$	2
	$1.269 < D' \leq 2.129$	$1.1628(2.129 - D') + 2$
	$1.129 < D' \leq 1.269$	3
	$0.269 < D' \leq 1.129$	$1.1628(1.129 - D') + 3$
	$D' \leq 0.269$	4

For the two safety sub-indexes ( $I_{FL}$  and  $I_{EX}$ ) and  $I_{MS}$ , they are each assessed by two properties, which are predicted using different prediction models. Each model has its distinct statistical performance indicator values and it is relatively hard to address to different uncertainties originated from two different prediction models. Thus, the initial form of  $I_{FL}$ ,  $I_{EX}$  and  $I_{MS}$  sub-indexes given by NFPA flammability rating, ISI, and IOHI respectively are applied, which are given in Tables 4.2, 4.3, and 4.6. Meanwhile, uncertainty is incorporated in the sub-indexes of  $I_{\eta}$ ,  $I_V$ ,  $I_{EL}$ , and  $I_{AH}$ , in which their revised scores are provided in Tables 6.3 to 6.6.

#### 6.4.3.3 Allocation of Sub-index Scores with Disjunctive Programming

In this stage, disjunctive programming algorithm presented in Section 6.3.2 are applied to assignment scores of the four revised sub-

indexes that account for property prediction uncertainty (as shown in Tables 6.3 to 6.6), as multiple score functions are involved in these sub-indexes. Meanwhile, the simpler disjunctive programming procedure as illustrated in Section 5.3.3.2 are utilised for the assignment of scores to the remaining three sub-indexes ( $I_F$ ,  $I_{EX}$  and  $I_{MS}$ ) that do not incorporate uncertainty.

#### 6.4.4 Molecular Design

In this case study, only first-order groups are employed to construct the molecules. The appropriate molecular building blocks selected to construct the solvent are based on the molecular structures of the conventionally used solvent for the extraction of carotenoids. These solvents include straight chain hydrocarbons, alcohols and ketones (Ibrahim and Onwuala, 2007). Besides, Yara-Varón et al. (2016a) proposed five green solvents namely 2-methyltetrahydrofuran (2-MeTHF), dimethyl carbonate (DMC), cyclopentyl methyl ether (CPME), isopropyl alcohol (IPA), and ethyl acetate, for the substitution of *n*-hexane to extract carotenoids from carrots. Based on the proposed solvents, this case study considers both acyclic and monocyclic compounds, and the selected molecular blocks include CH<sub>3</sub>, CH<sub>2</sub>, CH, C, OH, CH<sub>3</sub>CO, CH<sub>2</sub>CO, CH<sub>3</sub>O, CH<sub>2</sub>O, CHO, CH<sub>3</sub>COO, CH<sub>2</sub>COO, CH<sub>2</sub> (cyclic), CH (cyclic), C (cyclic), and O (cyclic). The structural constraints for first-order groups as presented in Section 5.3.4.1 are employed.

#### 6.4.5 Optimisation Model

In this final stage, the four selected design objectives are optimised. The first objective is to minimise the boiling point difference ( $T_{b,diff}$ ) between hexane and the solvent. The  $T_b$  of the solvent is first transformed

in terms of property operator,  $\Omega_{T_b}$ , which is given by the  $f(P)$  function of its GC model as shown in Table 4.12. For hexane, its  $T_b$  is 68.5°C or 341.65 K, where its corresponding property operator,  $\Omega_{T_b, \text{hexane}}$  is equivalent to 4.0441. Therefore for this first objective as shown in Equation (6.7), the absolute difference ( $\Omega_{T_b, \text{diff}}$ ) between the  $T_b$  property operator of solvent,  $\Omega_{T_b}$  and that of hexane,  $\Omega_{T_b, \text{hexane}}$  is minimised.

$$\min \Omega_{T_b, \text{diff}} = |\Omega_{T_b} - \Omega_{T_b, \text{hexane}}| = |\Omega_{T_b} - 4.0441| \quad (6.7)$$

For the second design objective, the heat of vaporisation of the solvent,  $H_v$  is minimised. Similarly to  $T_b$ , the  $H_v$  of solvent is also transformed into its corresponding property operator,  $\Omega_{H_v}$ , which is also given in Table 4.12. As for the third design objective, the solubility of carotenoids in the solvent is maximised for higher carotenoid extraction. HSP is utilised to predict the solubility of carotenoids in the solvent. There are three parameters in HSP, namely  $\delta_d$ ,  $\delta_p$ , and  $\delta_h$ , which signify the dispersion, polar, and hydrogen bonding respectively. Another parameter known as the distance of a solvent from the centre of the Hansen solubility sphere,  $R_a$  is calculated by the following equation.

$$R_a^2 = 4(\delta_{d,A} - \delta_{d,B})^2 + (\delta_{p,A} - \delta_{p,B})^2 + (\delta_{h,A} - \delta_{h,B})^2 \quad (6.8)$$

In Equation (6.8), component A refers to the solute (carotenoids) while component B refers to the solvent. The smaller the  $R_a$ , the greater the affinity between carotenoids and solvent B; thus the higher the solubility of carotenoids in the solvent. According to Choo et al. (1996), the three main constituents of carotenoids found in the residual oil in PPF are  $\beta$ -carotene (31.0%),  $\alpha$ -carotene (19.5%), and lycopene (14.1%). The HSP

values for these constituents can be taken from Aissou et al. (2017). Their HSP values are relatively close to one another. In this work, the HSP values used for carotenoids are the weighted average of the HSP values from the three constituents, in which the calculated  $\delta_d$ ,  $\delta_p$ , and  $\delta_h$  are 17.3782, 0.3839, and 1.6396 respectively. With these three parameters identified, the third design objective is to minimise the value of  $R_a$  for a high extraction performance. As for the fourth design objective,  $I_{SHI}$  is minimised for the synthesis of solvent with low hazard level. Unlike  $T_b$  and  $H_v$ , both  $R_a$  and  $I_{SHI}$  are not transformed into property operators as the latter two are not directly calculated by any prediction models. In summary, the objective functions of this case study are to minimise  $\Omega_{Tb,diff}$ ,  $\Omega_{Hv}$ ,  $R_a$ , and  $I_{SHI}$ .

In order to minimise the four design objective properties, the linear membership functions as given by Equations (4.5) and (5.19) are applied. The next step is to identify the upper and lower bounds for the four target properties. Each of these four properties is optimised one at a time to identify the property bounds for all four properties. Once all boundary values are identified, the four objectives are then expressed by the linear membership functions as given by Equations (6.9) to (6.12):

$$\frac{2.7261 - \Omega_{Tb,diff}}{2.7261 - 0.0032} = \lambda_{Tb,diff} \quad (6.9)$$

$$\frac{30.8018 - \Omega_{Hv}}{30.8018 - 14.0244} = \lambda_{Hv} \quad (6.10)$$

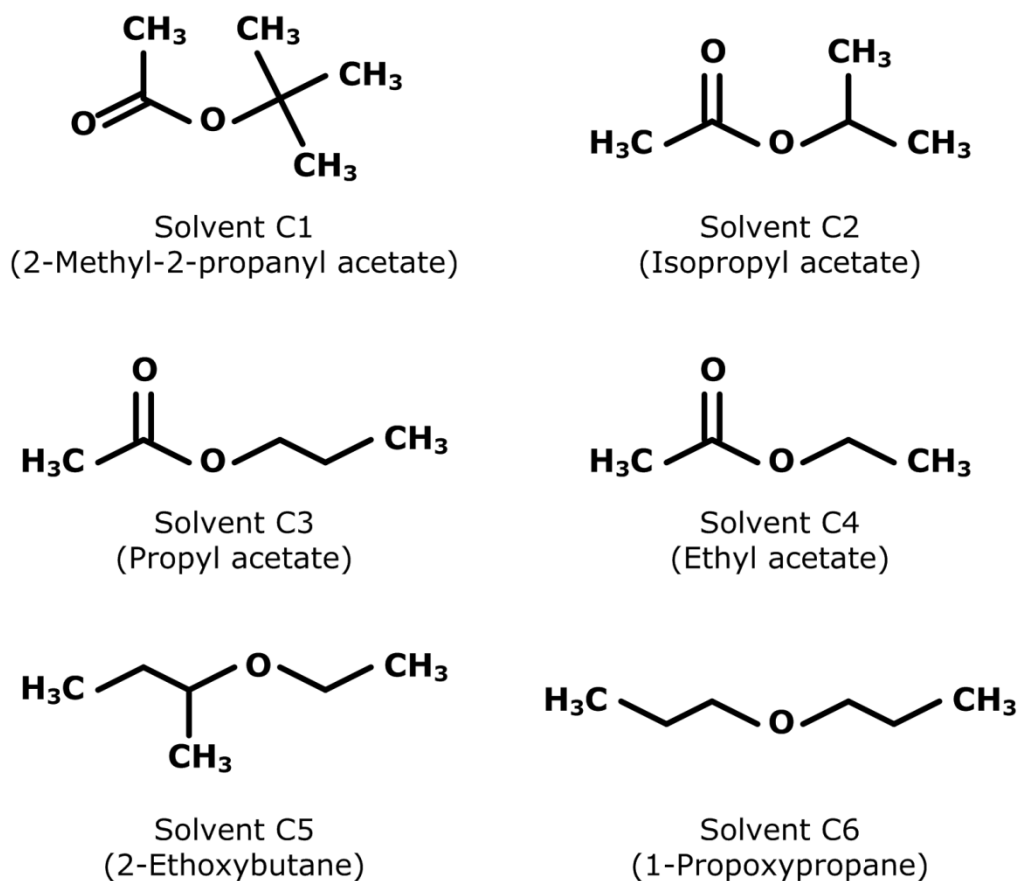
$$\frac{9.777867 - R_a}{9.777867 - 3.586346} = \lambda_{Ra} \quad (6.11)$$

$$\frac{11.80962 - I_{SHI}}{11.80962 - 8.767422} = \lambda_{ISHI} \quad (6.12)$$

The overall objective function of the optimisation model is given by Equation (5.20), subjected to constraints (4.8) and (5.21). Integer cuts have been used to generate alternate solutions so that there are multiple solutions to conduct experimental verification and to choose the final product.

#### 6.4.6 Results and Discussions

The optimisation model is solved using LINGO 14.0 with a computational time of 35 seconds for the first generated solution. From the optimisation results, the six molecules with the six highest  $\lambda$  values are shown in Figure 6.5.



**Figure 6.5:** The generated solvents with their molecular structures

**Table 6.7:** The six generated solvents with their properties

Solvent $\lambda$	$I_{SHI}$	$R_a$	$H_v$	$T_b$	$F_p$	$\log K_{ow}LC_{50}$	
			(kJ/mol)	(°C)	(°C)	(mg/l)	
C1	0.404	10.58	6.634	35.52	107.2	16.7	1.561 68.6
C2	0.379	10.30	7.432	34.32	86.2	9.6	1.011 114.6
C3	0.363	10.34	7.532	35.23	101.5	16.0	1.197 86.3
C4	0.326	10.24	7.761	32.89	68.7	4.5	0.762 208.0
C5	0.215	11.15	6.271	30.76	79.7	2.1	2.064 224.9
C6	0.204	11.19	6.327	31.67	95.4	8.5	2.250 169.3

**Table 6.8:** The six generated solvents with their sub-index scores

Solvent	$I_{FL}$	$I_{EX}$	$I_\eta$	$I_{MS}$	$I_V$	$I_{EL}$	$I_{AH}$	$I_{SHI}$
C1	3	1	1.35	2	1	1.40	0.83	10.58
C2	3	1	1.20	2	1	1.33	0.77	10.30
C3	3	1	1.24	2	1	1.36	0.74	10.34
C4	3	1	1.18	2	1	1.27	0.79	10.24
C5	3	1	1.16	2	1	1.71	1.28	11.15
C6	3	1	1.19	2	1	1.75	1.25	11.19

The estimated properties of these six solvents are shown in Table 6.7, while their respective sub-index scores are illustrated in Table 6.8. From the results, solvent C1 is the best optimal solution, as it has a reasonable  $I_{SHI}$  and  $R_a$ . However, it has the highest  $H_v$  and  $T_b$  among six solvents. By comparing the four optimised target properties individually, solvent C5 has the lowest  $R_a$  and  $H_v$ , but displays a relatively high  $I_{SHI}$ . Meanwhile, solvent C4 has the lowest  $I_{SHI}$  and  $T_b$ , but exhibits the largest  $R_a$ . Even though solvent C4 (ethyl acetate) has the largest  $R_a$ , Yara-Varón et al. (2016a) found that it demonstrates similar carotenoid extraction yield



as compared to that of hexane. Therefore, the other five solvents with lower  $R_a$  than solvent C4 should also possess compatible or better extraction yield as compared to hexane. In order to guide a decision maker for the selection of final candidate substance, one can opt for solvent C4 in the case where safety and health aspects are given the highest priorities, or solvent C5 for the extraction process to achieve the best performance.

When comparing the individual sub-index scores in Table 6.8, all solvents have a similar  $I_{FL}$  value of three, which is considered easily flammable. But these six solvents have relatively higher  $F_p$  compared to hexane, with solvent C1 offering the highest  $F_p$ . Since all six flash points fall in the same interval in the  $I_{FL}$  sub-index, they exhibit the same  $I_{FL}$  score. As for  $I_{EX}$ , they have a similar penalty of one, which indicates a very low tendency to cause explosion. The six solvents also have fairly similar  $\eta$ , since their  $I_\eta$  scores do not vary significantly. Meanwhile, all solvents exist as liquid form in standard condition, thus they display the same score for material phase sub-index  $I_{MS}$ . They also have low volatility, as all their  $I_V$  scores are one. As for  $I_{EL}$  and  $I_{AH}$ , in which the former measures chronic toxicity (inhalation) while the latter assesses acute toxicity (oral), solvents C1 to C4 have considerably low toxicity compared to solvents C5 and C6. Solvent C4 is considered the best choice of solvent if the aspects of safety and health were the only priorities of the design. If uncertainty analysis were not conducted on the sub-indexes, the initial forms of sub-indexes would be applied and solvents C1 to C4 would have a similar  $I_{SHI}$  value of 10 while solvents C5 and C6 would have an  $I_{SHI}$  score of 11. In this scenario, it is easy to conclude that solvents C5 and C6 are more hazardous than the remaining four solvents. However, the remaining four solvents display the similar  $I_{SHI}$  value, thus it is relatively difficult to differentiate their inherent hazard level. Hence, by carrying out uncertainty

analysis on the sub-indexes, it definitely helps to produce a total index score which better represents the actual inherent safety and health hazard level of the molecule.

By comparing the three case studies carried out as shown in Sections 4.4, 5.4 and 6.4, some safety and health sub-indexes have higher importance level towards the overall molecular hazard level while others contributed less. For instance, both  $I_{\eta}$  and  $I_{EL}$  have higher scorings as compared to the other sub-index. Hence, the two aforementioned sub-indexes have greater impact to the overall hazard in the case studies conducted in Sections 4.4 and 5.4, while the sub-index with highest hazard contribution in this case study is  $I_{FL}$ . The higher-impact (larger-scoring) sub-indexes are different among the case studies as they are dependent on the nature of the design problems. The current approach used to quantify the overall inherent hazard level is by summing up all the seven sub-indexes. With this approach, all sub-indexes are treated equally as weight factors are not introduced to the sub-indexes. However, this may not be the case as all sub-indexes contributed differently as demonstrated by the results of the three case studies. The hazard quantification method can be improved by assigning a larger weight to sub-index with higher impact, and vice versa. This new method prioritises sub-indexes with higher severity penalty scores to ensure that the model can formulate a more conservative solution with respect to safety and health hazard. Chapter 7 will study the effect of introducing weight factors to the sub-indexes on the optimal solutions generated by the model. Besides, the algorithm of AHP will also be illustrated to identify the appropriate weight factors.

## 6.5 Summary

A chemical product design framework employing CAMD methods has been applied to design a molecule with low safety and health hazards level that also meets a set of desired properties specified by the user. The existing safety and health indexes from literature are integrated into a CAMD problem to evaluate the safety and health characteristics of the generated molecules. In this work, the main highlight is the management of uncertainty on the sub-indexes to ensure that a more accurate index value is assigned to the molecule to better represent the safety and health performance of the molecule. The statistical performance indicators of the property prediction models are identified and applied to determine the uncertainty range in the sub-indexes. A case study on the solvent design for the extraction of carotenoids from PPF is carried out and fuzzy optimisation is applied to develop solvents that simultaneously achieve high product functionality and favourable safety and health performance. The results show that total penalty score,  $I_{SHI}$  of the molecules differ from one another. This helps to differentiate and compare the intrinsic hazard level exhibited by each molecule. Another way to improve the accuracy of the sub-index scores in quantifying the inherent hazard level of the molecules is to enhance the sensitivity of the scores. The property intervals can be smoothened to ensure that there is a continuous change of scorings from the lower bound to the upper bound of the feasible property value range. Besides, the current method of quantifying the intrinsic hazard level of the molecule is by summing up the seven sub-index values. Further work can be conducted to improve this approach, which should also take into account the severity of each sub-index score of the synthesised molecules.

## **CHAPTER 7**

# **ENHANCING MOLECULAR SAFETY AND HEALTH MEASUREMENT VIA INDEX SMOOTHING AND PRIORITISATION**

### **7.1 Introduction**

As presented in Chapters 5 and 6, a CAMD algorithm incorporating the inherent safety and occupational health sub-indexes have been presented to generate molecules that are optimised with respect to targeted functionalities, safety and health performance. In each sub-index, the evaluation is carried by one or more properties, in which they are divided into multiple sub-ranges, each of which is assigned a penalty score that corresponds to its degree of hazard. As the property value moves from one sub-range to another, the sub-index score switches abruptly at the property boundary, which is the point that separates two adjacent sub-ranges. This condition has created discontinuity in scores allocation, which is a major limitation in quantifying the molecular hazard level. In this chapter, the penalty scores are revised in a manner that there are smooth transitions from one property sub-range to another at the property boundaries. In addition, weight factors are introduced to the sub-indexes to ensure that a higher-impact sub-index is prioritised and hence assigned a larger weight. This approach can be carried out through the combination of the ordered weighted averaging (OWA) operator method with the analytic hierarchy process (AHP), which improves the quantification of the overall safety and health performance. In this chapter, AHP plays a role in

identifying the proper weight factors that will be introduced to the sub-indexes. This ensures that a higher-importance sub-index is emphasised through the allocation of larger weight.

## **7.2 Problem Statement**

The technique used to quantify the inherent risk level of a molecule as proposed in Chapters 4 to 6 is by summing up all the sub-index scores from the selected safety and health sub-indexes. This method presumes that all sub-indexes have equal impact to the plant safety and occupational health of the workers. However, from the results generated by the case studies in Sections 4.4, 5.4, and 6.4, certain sub-indexes have contributed more to the final total index score as compared to others. Besides, a molecule with the lowest total index score may not necessarily display low risk level with respect to all sub-indexes. This is because one of its sub-indexes may exhibit a very high score, while the remaining sub-indexes have relatively lower scores to compensate the high-scoring sub-index. In reality, any molecule demonstrating highly hazardous attribute with respect to a specific sub-index (eg. highly toxic) is usually deemed a dangerous chemical, and that particular molecule is therefore screened out in the chemical selection phase. Hence, there is a need to address this limitation on hazard quantification so that the more hazardous or greater-impact sub-index will be penalised to a greater extent. This limitation can be addressed by assigning different weight factors to each sub-index. One way to identify the weights is through AHP, which is a structured multi-attribute decision technique. Once AHP has identified the proper weights to be assigned to the sub-indexes, the next step is to consider utilising an approach that can assign weights in such a manner that sub-index with larger penalty score (higher adverse impact to the plant) receives a heavier

weight, and vice versa. OWA operator method can be adequately applied in this case, as it allows the allocation of weight factors based on the ordered ranking of the sub-index scores. It is also necessary to formulate the OWA operator as a mathematical model so that it can be integrated into a CAMD programming.

Besides, as shown by the general trend for the allocation of sub-indexes illustrated in Figure 6.1, the properties examined in the sub-indexes are divided into multiple sub-ranges, and numerical penalty scores are then assigned to each sub-range. Through this score allocation method, property values that fall in the same sub-range are deemed to possess the similar hazard level. The main limitation of this current allocation method is the discontinuity at the property boundaries, which may distort the comparison of alternatives that are near these limits. This weakness is addressed in this chapter by smoothing the scorings near the property boundaries. In summary, the issues to be addressed in this chapter are as following:

1. to determine weight factors given to the sub-indexes through AHP;
2. to assign weights to the sub-indexes in an ordered manner through OWA operator;
3. to represent OWA operator as a mathematical model to be incorporated into a CAMD;
4. to smoothen the sub-index scores at the property boundaries.

Section 7.3 illustrates the working procedure of an AHP, while the OWA operator method is presented in Section 7.4.

### 7.3 Analytic Hierarchy Process (AHP)

In many decision-making problems, AHP is often used as the decision structuring and analysis tool. The application of AHP has three main principles, namely problem decomposition, comparative judgements, and synthesis of priorities. The stage of decomposition begins by breaking down a problem into smaller elements, and structuring them into a hierarchy with different levels, where each level contains a finite number of elements. The top of the hierarchy is usually the objectives of the decision-making problem, the intermediate levels are represented by several criteria on which the subsequent levels depend on, while the lowest level contains a list of potential alternatives. Each element in the hierarchy will serve as the criterion for all elements of the level below (Saaty and Kearns, 1985).

In the stage of comparative judgements, pairwise comparisons of the relative strength or importance of  $n$  elements in the same hierarchy level are carried out with respect to a common aspect in the level above. Decision-maker can compare any two elements (eg.  $E_i$  and  $E_j$ ) and assign a numerical scale  $a_{ij}$  as the ratio of their relative importance. If two elements being compared have equal importance, then  $a_{ij}$  would be given a value one. If element  $E_i$  is considered to have higher importance than  $E_j$ , then  $a_{ij}$  would be greater than one. Once  $n(n-1)/2$  pairwise comparisons are completed among all elements, a positive reciprocal square matrix containing the comparative judgments is obtained as shown in Equation (7.1).

$$A = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & \dots & \dots & \dots \\ a_{n1} & a_{n2} & \dots & a_{nn} \end{bmatrix} \quad (7.1)$$

In the comparison matrix, the reciprocal property  $a_{ji} = 1/a_{ij}$  (where  $a_{ij} > 0$ ) always holds true for  $j = 1, 2, \dots, n$  and  $i = 1, 2, \dots, n$ , in which  $n$  represents the number of elements in the level. A measurement scale of 1 to 9 (Saaty, 1977) as shown in Table 7.1 is used to assign the numerical value to each  $a_{ij}$ . For instance, if element  $E_i$  is deemed to be strongly more important than  $E_j$ , then its  $a_{ij}$  value would be 5. Its corresponding reciprocal,  $a_{ji}$  will then be equivalent to  $1/5$ .

**Table 7.1:** The fundamental AHP scale (Saaty, 1977)

Numerical scale	Definition (explanation)
1	Equal importance (two elements contribute equally to the objective)
3	Moderate importance of one over another (experience and judgement slightly favour one element over another)
5	Essential or strong importance (experience and judgement strongly favour one element over another)
7	Very strong importance (an element is strongly favoured and its dominance demonstrated in practice)
9	Extreme importance (the evidence favouring one element over another is of the highest possible order of affirmation)
2, 4, 6 and 8	Intermediate values between the two adjacent judgements (the judgement falls between two levels)
Reciprocals	If element $i$ has one of the above numbers allocated to it when compared with element $j$ , then $j$ has the reciprocal value when compared with $i$

The relative strength or importance of all elements being compared can be identified from the comparison matrix. To do this, a vector of



priorities or weights,  $w$  has to be computed from the matrix. First, in order to determine vector  $w$ , Saaty (1980) has suggested the principal eigenvalue method as shown in Equation (7.2):

$$A \cdot w = \lambda_{\max} \cdot w \quad (7.2)$$

where  $A$  is the pairwise comparison matrix and  $\lambda_{\max}$  is the maximum eigenvalue of matrix  $A$ . The solution for vector  $w$  is determined numerically by raising the matrix  $A$  to a sufficiently large power, then summing over the rows and normalising them to obtain the weight vector  $w = (w_1, w_2, \dots, w_n)^T$ . The value of  $\lambda_{\max}$  can then be identified through Equation (7.2). The next step is to determine the consistency of the comparison matrix. A matrix is considered to be perfectly consistent when its elements fulfil the following condition:

$$a_{ij} \cdot a_{jk} = a_{ik} \quad \forall i, j, k \quad (7.3)$$

Given a problem with three elements where element  $E_i$  is considered to be more important than  $E_j$ , and  $E_j$  is more important than  $E_k$ . The problem is deemed to be inconsistent if element  $E_k$  has higher importance than  $E_i$ . Saaty (1980) has introduced Equations (7.4) and (7.5) to measure the extent of the deviation from consistency of the comparison matrix:

$$CI = \frac{\lambda_{\max} - n}{n - 1} \quad (7.4)$$

$$CR = \frac{CI}{RI} \quad (7.5)$$

where  $CR$  is the consistency ratio,  $CI$  is the consistency index, and  $RI$  is a random consistency index that can be referred to Table 7.2. The value of  $CR$  should be less than 0.1 or 10% for the deviation from consistency to be acceptable. If the calculated  $CR$  does not fall within this range, decision makers will be asked to revise their comparative judgements. Since the AHP numerical scale as shown in Table 7.1 ranges from one to nine, it is suggested that one should not consider more than seven elements for pairwise comparison in order to ensure the validity of numerical comparisons. In case of a problem with large number of elements, hierarchical decomposition should be carried out by grouping the elements into comparability classes of approximately seven elements each (Saaty and Kearns, 1985).

**Table 7.2:** Random consistency ( $RC$ ) index (Saaty, 1980)

Size of matrix, $n$	$RC$	Size of matrix, $n$	$RC$
1	0	6	1.24
2	0	7	1.32
3	0.58	8	1.41
4	0.90	9	1.45
5	1.12	10	1.49

In the synthesis of priorities stage, the priorities of all elements are determined for each level beginning from the second level to the bottom level. For each element, its composite or global priority is calculated by multiplying the priority of its respective criterion in the level above and summing them for each element in a level based on the criteria it influences (Saaty and Kearns, 1985). This stage is illustrated with an example of decision making problem that contains three levels, where the top level is represented by the objective. In its second level, there are two

criteria known as K1 and K2, in which the local priorities calculated by pairwise comparison with respect to the objective are 0.667 and 0.333 respectively. Meanwhile in the third level, each criterion is provided with two alternatives known as J1 and J2. Given the local priorities of J1 and J2 with respect to criterion K1 are 0.75 and 0.25 respectively, while the local priorities of J1 and J2 with respect to K2 are 0.2 and 0.8 respectively. Then, the global priority for J1 can be calculated by:  $(0.667 \times 0.75) + (0.333 \times 0.2) = 0.567$ , while the global priority for J2 is determined by:  $(0.667 \times 0.25) + (0.333 \times 0.8) = 0.433$ . Alternative J1 is chosen as the preferred solution as it has larger global priority than that of J2. The application of AHP to determine the weight factors is further illustrated in Section 7.5.2.

## **7.4 Ordered Weighted Averaging (OWA) Operator**

As AHP is applied to determine weights to the sub-indexes, the next step is to employ a method that can assign weights in such a manner that sub-index with higher penalty score receives a larger weight, and vice versa. This ensures that a higher-scoring sub-index will have a greater contribution to the final weighted  $I_{SHI}$  value. This results in an inherently conservative weighing procedure, which can be managed by applying aggregation operators that have been developed to assist in aggregating information. Some of these methods include the max and min operators, arithmetic averaging (AA) operator, weighted AA (WAA) operator, geometric averaging (GA) operator, ordered weighted averaging (OWA) operator, etc. (Xu and Da, 2003). In this work, the weights are assigned based on the ordered position of the sub-index scores. The allocation of weight in this manner can be adequately done with the application OWA

operator method introduced by Yager (1988). An OWA operator of dimension  $n$  is a function

$$F : R^n \rightarrow R$$

with an associated  $n$  vector

$$w = (w_1, w_2, \dots, w_n)^T$$

where

1.  $w_i \in [0,1]$
2.  $\sum_i w_i = 1$

Besides,

$$F(a_1, a_2, \dots, a_n) = \sum_j w_j b_j \quad (7.6)$$

where  $b_j$  is the  $j$ th largest of the  $a_i$ . The principal characteristic of the OWA operator is the reordering step, in which an argument  $a_i$  is not associated with a specific weight  $w_i$ , but a weight  $w_i$  is associated with a specific rank  $i$  of the arguments (Yager, 1988). The application of OWA operator approach in this chapter is further demonstrated in Section 7.5.2.

## 7.5 Methodology

As shown in Figure 3.3, the procedure proposed in this chapter is an extension of the methodology presented in Chapter 5, in which the goals are to introduce weight factors to the sub-indexes and to smoothen the sub-index scores at the property boundaries.

### 7.5.1 Problem Formulation and Safety and Health Assessment

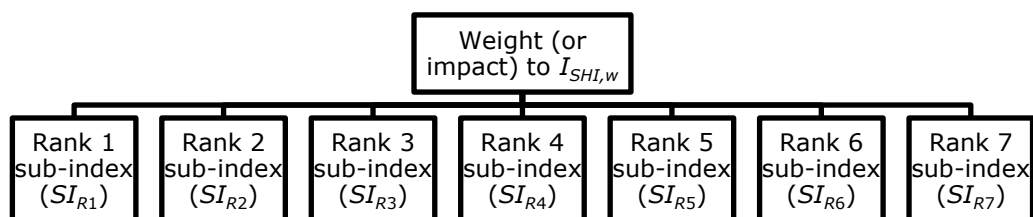
The first stage of this methodology is the problem formulation phase, which serves the same purpose as the step presented in Section 5.3.1. The following phase is the safety and health assessment step, which considers the same seven sub-indexes as discussed in Section 5.3.2 to evaluate the molecular performance.

### 7.5.2 Determination of Total Weighted Index Score

As presented in Section 5.3.2, the total index score,  $I_{SHI}$  assigned to the molecule is equivalent to the summation of all seven sub-index scores. A lower  $I_{SHI}$  is favourable as the molecule exhibits lower magnitude of hazard. This method of measuring hazard treats all sub-indexes with equal importance, as they all have the similar weight of contribution towards the final total score. In this chapter, weights are introduced to the sub-indexes and the quantification of the overall hazard exhibited by the molecules can be carried out through OWA operator approach, as shown by Equation (7.7). This equation is derived from Equation (7.6), where the right-hand side of the former equation is the summation for all multiplications between the sub-index values and their corresponding weights.  $SI_{R1}$  represents the sub-index with the highest scoring,  $SI_{R2}$  is the sub-index with the second highest scoring, and so on.  $w_1$  to  $w_7$  are the weights given to the sub-indexes, in which the value of  $w_1$  is greater than  $w_2$ ,  $w_2$  is larger than  $w_3$ , and so on. As weights are now introduced in the equation, the overall hazard level of the molecule is quantified by  $I_{SHI,w}$ , which signifies the total weighted index score.

$$I_{SHI,w} = w_1 SI_{R1} + w_2 SI_{R2} + w_3 SI_{R3} + \dots + w_6 SI_{R6} + w_7 SI_{R7} \quad (7.7)$$

Through OWA operator method, the weights can be assigned conditionally to the criteria depending, based on the logic that there should be more weight placed on the weaker and more critical features of a given alternative. This approach can thus yield a more conservative approach than simply assigning fixed weights to the criteria. For the measurement of safety and health aspects of the molecules, sub-index with weaker performance (or more severe score) is assigned higher weight to penalise the high hazard condition demonstrated by the molecules. The next step is to employ AHP for the identification of weights introduced to the sub-indexes. In the first stage of AHP, the problem is represented in the form of hierarchy that contains several elements. The hierarchy for this weight determination problem is shown in Figure 7.1.



**Figure 7.1:** Hierarchy for the weight determination of sub-indexes

In Figure 7.1, the top level is the goal of this AHP problem, which is to determine the weight of the sub-indexes. The seven sub-indexes in the bottom level are the elements which contribute to the final  $I_{SHI,w}$  value. In the next step, pairwise comparisons are considered among all seven sub-indexes to assess the relative importance and impact towards  $I_{SHI,w}$ . The complete pairwise comparison matrix is shown in Table 7.3.

**Table 7.3:** Pairwise comparison matrix

	$SI_{R1}$	$SI_{R2}$	$SI_{R3}$	$SI_{R4}$	$SI_{R5}$	$SI_{R6}$	$SI_{R7}$
$SI_{R1}$	1	2	3	4	5	6	7
$SI_{R2}$	1/2	1	2	3	4	5	6
$SI_{R3}$	1/3	1/2	1	2	3	4	5
$SI_{R4}$	1/4	1/3	1/2	1	2	3	4
$SI_{R5}$	1/5	1/4	1/3	1/2	1	2	3
$SI_{R6}$	1/6	1/5	1/4	1/3	1/2	1	2
$SI_{R7}$	1/7	1/6	1/5	1/4	1/3	1/2	1

The first row of the matrix in Table 7.3 shows the pairwise comparisons made between the highest-scoring sub-index with the other sub-indexes. This highest-scoring sub-index is deemed to have an intermediate level of “equal importance” and “moderate importance” over the second highest-scoring sub-index, thus the allocated numerical scale for this comparison is two. Meanwhile, the highest-scoring sub-index is considered to have a very strong importance over the lowest-scoring sub-index, thus a numerical value of seven is assigned to this comparison. The numerical values in the first column are equivalent to the reciprocal values of the first row respectively. The diagonal values in the matrix are equal to unity as the sub-index being evaluated is compared with itself.

In the synthesis of priorities stage, the vector of weights,  $w$  has to be determined from the comparison matrix in Table 7.3. Through the maximum eigenvalue method as given by Equation (7.2), the weight vector,  $w$  calculated is shown as below:

$$w = [0.3543 \ 0.2399 \ 0.1587 \ 0.1036 \ 0.0676 \ 0.0448 \ 0.0312]^T \quad (7.8)$$

Hence, the values of  $w_1$  to  $w_7$  are 0.3543, 0.2399, 0.1587, 0.1036, 0.0676, 0.0448 and 0.0312 respectively. By substituting the vector  $w$  into Equation (7.2), the value of  $\lambda_{\max}$  determined from the aforementioned equation is 7.1955. The next step is to identify the extent of the deviation from consistency of the comparison matrix. By substituting the  $\lambda_{\max}$  value into Equation (7.4), the calculated value of  $CI$  is 0.0326. According to Table 7.2, the value of  $RI$  is 1.32 when the number of elements,  $n$  is 7. Therefore, the value of  $CR$  determined from Equation (7.5) is 0.0247 or 2.47%, which is lower than the 10% tolerance. Thus, the intensity of inconsistency of the matrix is acceptable.

In CAMD programming, constraints have to be introduced to allocate the descending order of sub-index values for  $SI_{R1}$  to  $SI_{R7}$  given in Equation (7.7). Binary integer variables are used to formulate the constraints, which are given by Equations (7.9) to (7.12):

$$b_{i1}I_{FL} + b_{i2}I_{EX} + b_{i3}I_{\eta} + b_{i4}I_{MS} + b_{i5}I_V + b_{i6}I_{EL} + b_{i7}I_{AH} = SI_{Ri} \quad i = 1 \dots 7 \quad (7.9)$$

$$b_{i1} + b_{i2} + b_{i3} + b_{i4} + b_{i5} + b_{i6} + b_{i7} = 1 \quad i = 1 \dots 7 \quad (7.10)$$

$$b_{1i} + b_{2i} + b_{3i} + b_{4i} + b_{5i} + b_{6i} + b_{7i} = 1 \quad i = 1 \dots 7 \quad (7.11)$$

$$SI_{Ri} \geq SI_{Rj} \quad i = 1 \dots 6, j = i + 1 \quad (7.12)$$

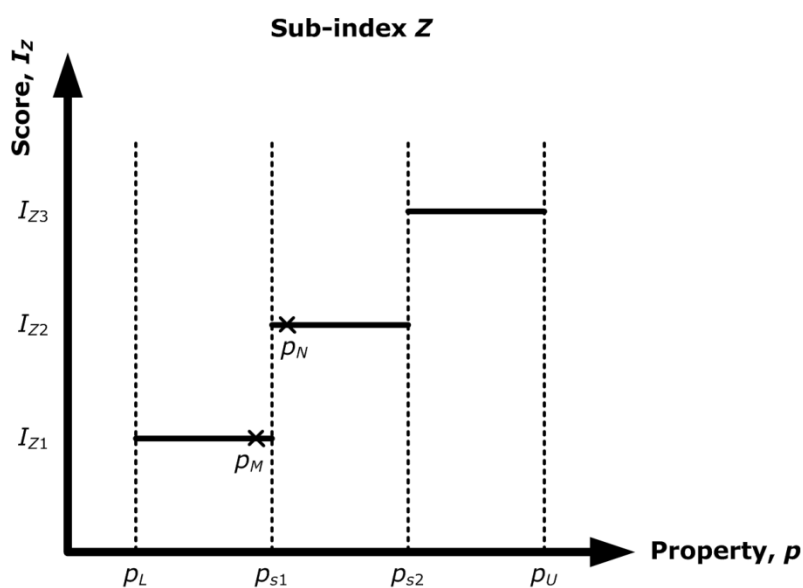
where  $b_{ij}$  (for  $i = 1 \dots 7$  and  $j = 1 \dots 7$ ) is the binary integer variable. With this new approach to quantify the molecular hazard, it can be noted that molecule with multiple severe-scoring sub-indexes will have considerably large  $I_{SHI,w}$  value. Therefore, the introduction of weights using OWA operator helps to enhance the final  $I_{SHI,w}$  score to accurately represent the overall intrinsic hazard level of a molecule.



The next stage is to identify the property prediction methods for all properties involved in the design problem. The GC models for properties assessed in the sub-indexes are provided in Tables 4.11, 4.14, and 6.2.

### 7.5.3 Smoothing Sub-index Scores

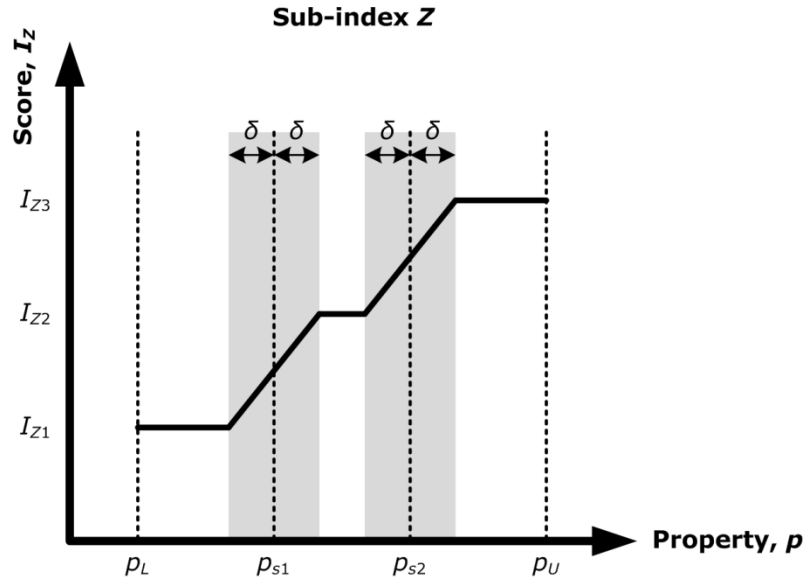
Based on the established inherent safety and health indexes, the allocation of sub-index scores is dependent on subjective scaling and weighting. The physicochemical properties assessed in the sub-indexes are divided into subjective ranges, where each range is then introduced a score depending on the authors' judgement. As mentioned in Section 7.2, the main limitation of this method is the discontinuity at the property boundary, which is illustrated by the following scenario in Figure 7.2.



**Figure 7.2:** A scenario of two property values with different sub-index scores

Given that there are two molecules with property values of  $p_M$  and  $p_N$  based on Figure 7.2.  $p_M$  falls near the upper bound of the " $p_L$  to  $p_{s1}$  range," which gives it a score of  $I_{z1}$ ; while  $p_N$  is close to the lower bound of

the " $p_{s1} - p_{s2}$  range," which returns a score of  $I_{Z2}$ . Both property values receive different sub-index scores as they fall in two sub-ranges with different hazard level. However, according to Figure 7.2, both  $p_M$  and  $p_N$  are close to one another, so it is possible that they might exhibit a similar hazard level. Therefore, the comparison of two property values near to the same property boundary but locating at different sub-ranges is not efficient as there exists discontinuity at the property boundary where the score switches abruptly from one value to another. This issue is addressed in this chapter, where the sub-index scores at the property boundary region will be smoothed to ensure that there is continuity for the allocation of sub-index scores at any property value. The resulted general trend for smoothed sub-index scores is shown in Figure 7.3 This modification can enhance the comparison of hazard level between the two property values with similar condition as the scenario in Figure 7.2.



**Figure 7.3:** Revised sub-index with smoothed scores

In Figure 7.3, the grey regions represent the smoothed regions where linear slopes are introduced to transit the scores from one sub-index

value to another. The wideness of the smoothened region is determined by the value of  $\delta$ . This value will identify the lower and upper bounds of the smoothened region for each property boundary. All smoothened regions should possess similar  $\delta$  value for consistency purpose. To determine  $\delta$ , a 10% margin of the property boundary values is applied. In general, a design factor of 10% is used for process flows to allow some flexibility in process operation (Sinnott, 2005). Therefore, a 10% safety/health factor is calculated for all property boundaries in Figure 7.3. Let us assume that the safety/health factors for  $p_{s1}$  and  $p_{s2}$  are  $\delta_1$  and  $\delta_2$  respectively. From Figure 7.3,  $p_{s2}$  has a larger numerical value than  $p_{s1}$ , so  $\delta_2$  is larger than  $\delta_1$ . As mentioned previously, a similar  $\delta$  value must be applied for all property boundaries. In order to ensure that all property boundaries can attain a minimum 10% factor, the selected  $\delta$  value should take the largest value among  $\delta_1$  and  $\delta_2$ . In this case,  $\delta$  is equivalent to  $\delta_2$ . In other words, the  $\delta$  value applied for a particular sub-index is the 10% factor of the largest property boundary (numerical) value.

#### 7.5.4 Allocation of Sub-index Scores with Disjunctive Programming

In this chapter, the algorithm to assign sub-index score to the molecule depending on its property value is developed. Based on the modified sub-index in Figure 7.3, the assigned sub-index score,  $I_z$  is given by:

$$I_z = \begin{cases} I_{z1} & p_L \leq p \leq p_{s1} - \delta \\ \frac{p - (p_{s1} - \delta)}{2\delta} (I_{z2} - I_{z1}) + I_{z1} & p_{s1} - \delta \leq p \leq p_{s1} + \delta \\ I_{z2} & p_{s1} + \delta \leq p \leq p_{s2} - \delta \\ \frac{p - (p_{s2} - \delta)}{2\delta} (I_{z3} - I_{z2}) + I_{z2} & p_{s2} - \delta \leq p \leq p_{s2} + \delta \\ I_{z3} & p_{s2} + \delta \leq p \leq p_U \end{cases} \quad (7.13)$$

Based on Equation (7.13), the property value,  $p$  is divided into five intervals. The sub-index score in each interval is defined by a distinct function. Thus, the sub-index score from the lower property value bound,  $p_L$  to the upper property value bound,  $p_U$  is never continuous. Disjunctive programming as discussed in Section 6.3.2 can be employed to formulate these discontinuous functions. The  $I_Z$  function in Equation (7.13) is transformed to the following mixed-integer formulation using four binary integer variables ( $b_p$ ,  $b_q$ ,  $b_r$  and  $b_s$ ):

$$I_Z = b_p I_{Z1} + (b_q - b_p) \left[ \frac{p - (p_{s1} - \delta)}{2\delta} (I_{Z2} - I_{Z1}) + I_{Z1} \right] + (b_r - b_q) I_{Z2} \\ + (b_s - b_r) \left[ \frac{p - (p_{s2} - \delta)}{2\delta} (I_{Z3} - I_{Z2}) + I_{Z2} \right] + (1 - b_s) I_{Z3} \quad (7.14)$$

where the binary integer variables are subjected to the following criteria:

$$b_p = \begin{cases} 0 & p \geq p_{s1} - \delta \\ 1 & p < p_{s1} - \delta \end{cases} \quad (7.15)$$

$$b_q = \begin{cases} 0 & p \geq p_{s1} + \delta \\ 1 & p < p_{s1} + \delta \end{cases} \quad (7.16)$$

$$b_r = \begin{cases} 0 & p \geq p_{s2} - \delta \\ 1 & p < p_{s2} - \delta \end{cases} \quad (7.17)$$

$$b_s = \begin{cases} 0 & p \geq p_{s2} + \delta \\ 1 & p < p_{s2} + \delta \end{cases} \quad (7.18)$$

The binary integer variables in Equation (7.14) act as switches to only 'activate' a single score function at a time to determine the sub-index score. To ensure that the model allocates the correct value to the binary integer variables in order to satisfy criteria (7.15) to (7.18), the following constraints are also imposed:

$$[p_L - (p_{s1} - \delta)]b_p < p - (p_{s1} - \delta) \leq [p_U - (p_{s1} - \delta)](1 - b_p) \quad (7.19)$$

$$[p_L - (p_{s1} + \delta)]b_q < p - (p_{s1} + \delta) \leq [p_U - (p_{s1} + \delta)](1 - b_q) \quad (7.20)$$

$$[p_L - (p_{s2} - \delta)]b_r < p - (p_{s2} - \delta) \leq [p_U - (p_{s2} - \delta)](1 - b_r) \quad (7.21)$$

$$[p_L - (p_{s2} + \delta)]b_s < p - (p_{s2} + \delta) \leq [p_U - (p_{s2} + \delta)](1 - b_s) \quad (7.22)$$

## 7.5.5 Molecular Design and Optimisation Model Formulation

In molecular design phase, the suitable building groups to form the potential molecular candidates are selected. Additional structural constraints on the molecular groups are implemented to ensure that structurally feasible molecules are generated. Only first-order molecular groups are considered in order to reduce the complexity of the CAMD programming, as the incorporation of disjunctive algorithm has resulted in several non-linear constraints [eg. Equation (7.14)]. As for the phase of optimisation model formulation, all target properties chosen as the design objectives are transformed into the linear membership functions. All objectives are simultaneously optimised to generate the optimal molecular structure that achieves the design goal. The complete procedures for both phases have been presented in Sections 5.3.4.1 and 5.3.5.

## 7.6 Case Study: Solvent Design for Extraction of Carotenoids

### 7.6.1 Problem Formulation

The same case study conducted in Section 6.4 is applied to demonstrate the methodology proposed in this chapter. The objective of the case study is to identify a solvent that can replace hexane to extract carotenoids from the residual oil found in palm pressed fibre (PPF). The

solvent should be able to lower the energy requirement needed for evaporation. Thus, it must have low boiling point ( $T_b$ ) and low heat of vaporisation ( $H_v$ ) so less energy is needed to heat up the solvent to its  $T_b$  and subsequently vaporise it. Besides, carotenoids should have high solubility in the developed solvent. The Hansen Solubility Parameters (HSP) are utilised to calculate the solubility of carotenoids in the solvent. The solubility is expressed by a parameter known as the distance of a solvent from the centre of the Hansen solubility sphere,  $R_a$  which can be calculated by Equation (6.8). The solvent must also achieve the desirable attributes for its safety and health aspects, such as low flammability and toxicity. This can be measured by the selected inherent safety and health sub-indexes. The total weighted index score,  $I_{SHI,w}$  in Equation (7.7) will be used to determine the inherent hazard level posed by the solvent. A low  $I_{SHI,w}$  score is preferred for an inherently safer and healthier solvent.

Overall, the four main objective functions in this case study is to minimise  $T_b$ , minimise  $H_v$ , minimise  $R_a$ , and minimise  $I_{SHI,w}$ . Since the conventionally used hexane is highly flammable and toxic to aquatic life, the flash point ( $F_p$ ) and acute toxicity (96-h  $LC_{50}$  to fathead minnow) are selected as property constraints. The lower bound of  $F_p$  is set at  $-13.3^{\circ}\text{C}$ , which is  $10^{\circ}\text{C}$  higher than that of hexane ( $-23.3^{\circ}\text{C}$ ). As for acute toxicity  $LC_{50}$ , the lower bound is set at 100 mg/l, which indicates that the solvent is not harmful to the aquatic environment as defined by the United Nations' Globally Harmonised System of Classification and Labelling of Chemicals (GHS).

For all the listed target properties, their corresponding property estimation models must be identified. The involved GC models have been provided in Section 6.4.3.1.

### 7.6.2 Safety and Health Assessment

The safety and health assessment on the solvent are measured using the selected safety and health sub-indexes that include flammability ( $I_{FL}$ ), explosiveness ( $I_{EX}$ ), viscosity ( $I_{\eta}$ ), material phase ( $I_{MS}$ ), volatility ( $I_V$ ), exposure limit ( $I_{EL}$ ), and acute health hazard ( $I_{AH}$ ). When developing the sub-indexes, the range of the sub-index scores is assigned based on the importance of the specific sub-index to the plant safety (Heikkilä, 1999) and the magnitude of impacts resulted by chemical exposure (Hassim and Hurme, 2010a). A more significant sub-index is allocated a larger scoring range. Four sub-indexes ( $I_{FL}$ ,  $I_{EX}$ ,  $I_{EL}$  and  $I_{AH}$ ) contain the largest scoring range of four, while two sub-indexes ( $I_{\eta}$  and  $I_{MS}$ ) display the smallest scoring range of two. It can be noticed that both flammability and explosiveness are the crucial parameters affecting the safety factors of the chemicals. As for occupational health, the parameters of exposure limit and acute health hazard, which both measure the toxicity of the chemical, cause more significant health impacts to the plant workers.

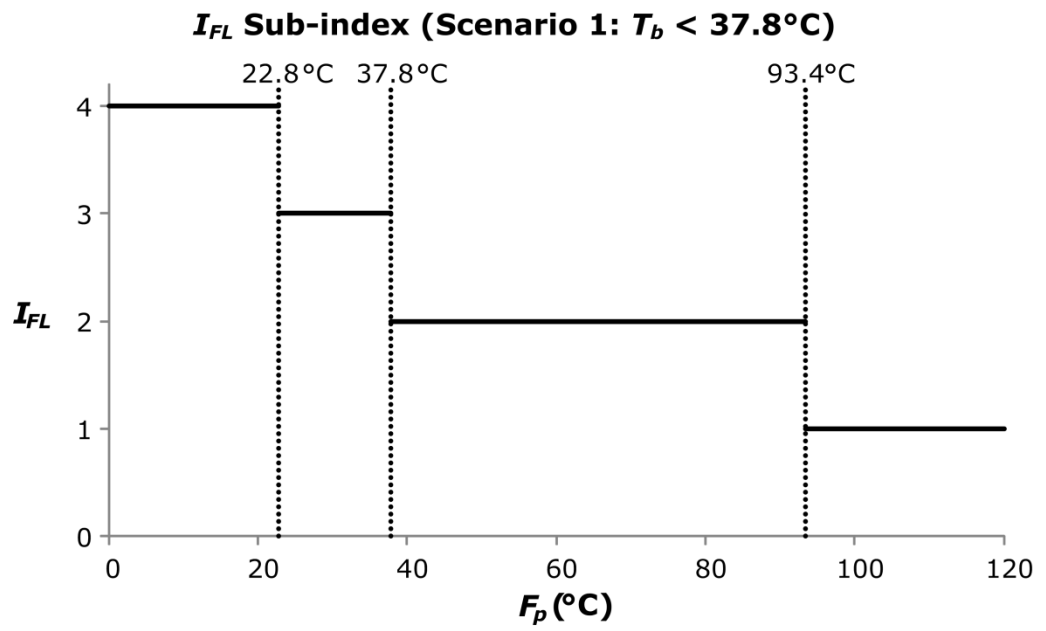
Among the selected sub-indexes, their lowest scores vary from zero to one while their highest scores differ from three to four. The scorings for all sub-indexes can be considered consistent as there are no sub-indexes with relatively high or low scores. In this work, the applied sub-index scores are in their initial forms. No scores normalisation is carried out on the seven sub-indexes, as such step causes all sub-indexes to be considered as equal importance. Equation (7.7) is applied to calculate  $I_{SHI,w}$ , where the seven sub-index scores are first sorted in descending order. The largest score is then multiplied with the largest weight, the second largest score is multiplied with the second largest weight, and so on. The proposed method to quantify the overall hazard of solvent ensures

that a particular sub-index displaying higher hazard is penalised more, so that it has larger impact on the  $I_{SHI,w}$  score.

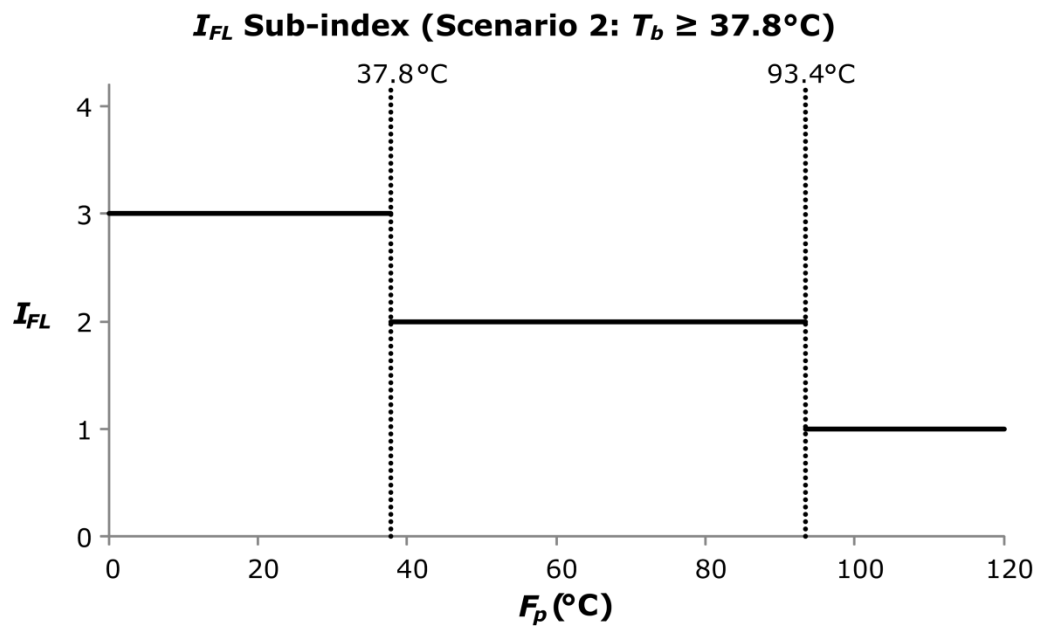
### 7.6.3 Smoothing Sub-index Scores

In this section, the smoothing of flammability sub-index ( $I_{FL}$ ) is demonstrated.  $I_{FL}$  evaluates the tendency of a material to burning in air. The allocation of  $I_{FL}$  scores are shown in Table 4.2, where the two properties assessed by this sub-index include flash point ( $F_p$ ) and boiling point ( $T_b$ ). There are three boundary values for  $F_p$  (93.4°C, 37.8°C and 22.8°C), while only a single boundary value for  $T_b$  (37.8°C). Since there is only one  $T_b$  boundary, the  $I_{FL}$  sub-index scores can be presented under two  $T_b$  scenarios, where the graphical representations are shown in Figure 7.4. Figure 7.4 shows the graphical illustration for the  $I_{FL}$  sub-index under two  $T_b$  conditions. For both scenarios, the sub-index scores are similar above  $F_p$  of 22.8°C. However, the scores differ below  $F_p$  of 22.8°C, in which the scores increase to four in Scenario 1 ( $T_b < 37.8^\circ\text{C}$ ), while remains at three in Scenario 2 ( $T_b \geq 37.8^\circ\text{C}$ ). Thus, there are only two  $F_p$  boundaries in Scenario 2, namely  $F_p$  at 37.8°C and 22.8°C. At all  $F_p$  boundaries, the  $I_{FL}$  scores switch abruptly from one value to another. In this section, the scores at the boundary regions will be smoothened to address the discontinuity issue mentioned in Section 7.2. The entire smoothing is carried out in two parts; the first part is conducted in terms of  $F_p$ , while the next part is in terms of  $T_b$ . As mentioned previously, the  $F_p$  boundaries are 93.4°C, 37.8°C, and 22.8°C. The 10% margin of the three boundary values are 9.34°C, 3.78°C and 2.28°C respectively. From the three margin values, the largest value, 9.34°C is chosen as the overall  $F_p$  margin. The selection



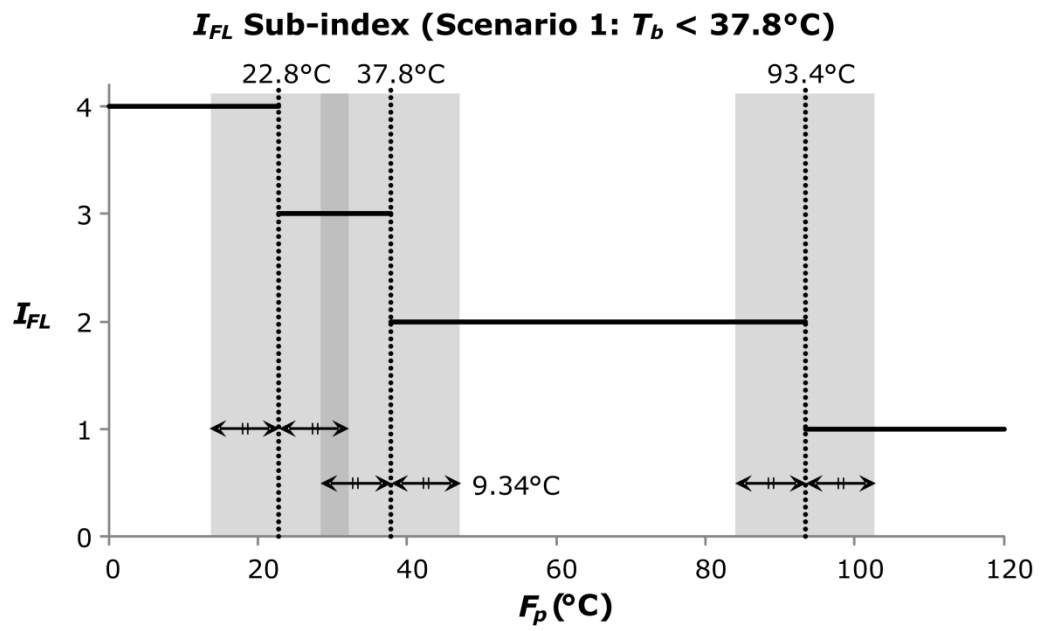


(a)

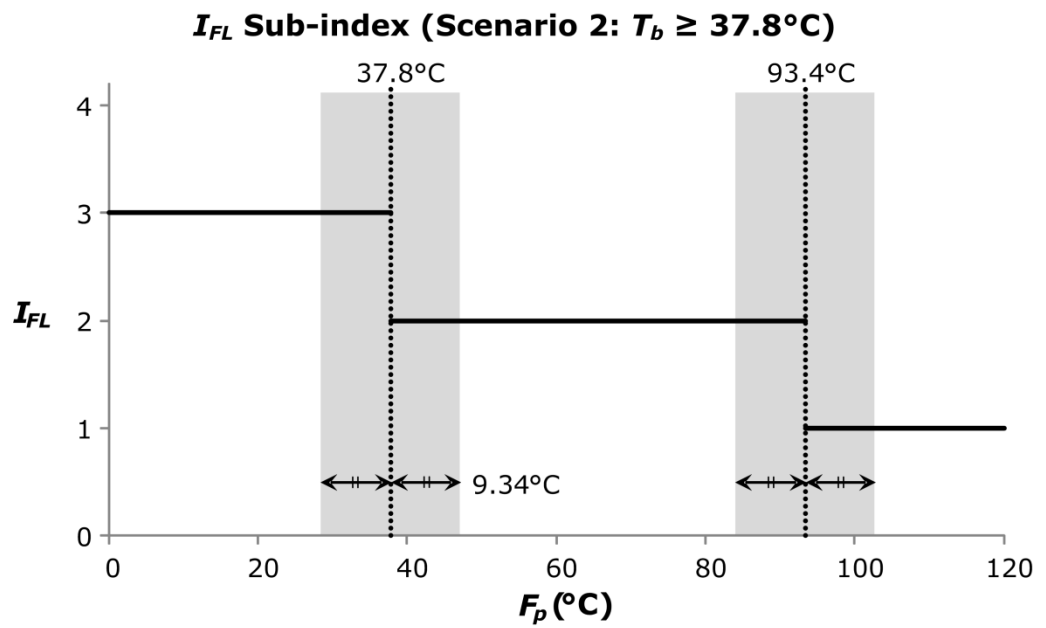


(b)

**Figure 7.4:** Sub-index scores of  $I_{FL}$  when (a)  $T_b < 37.8^\circ\text{C}$ ; (b)  $T_b \geq 37.8^\circ\text{C}$

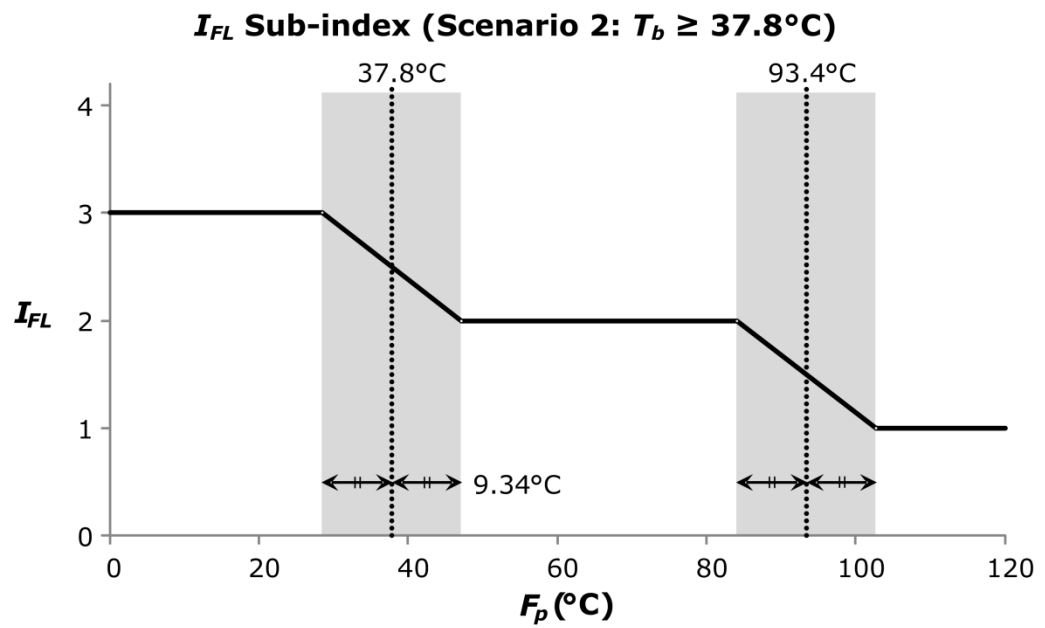
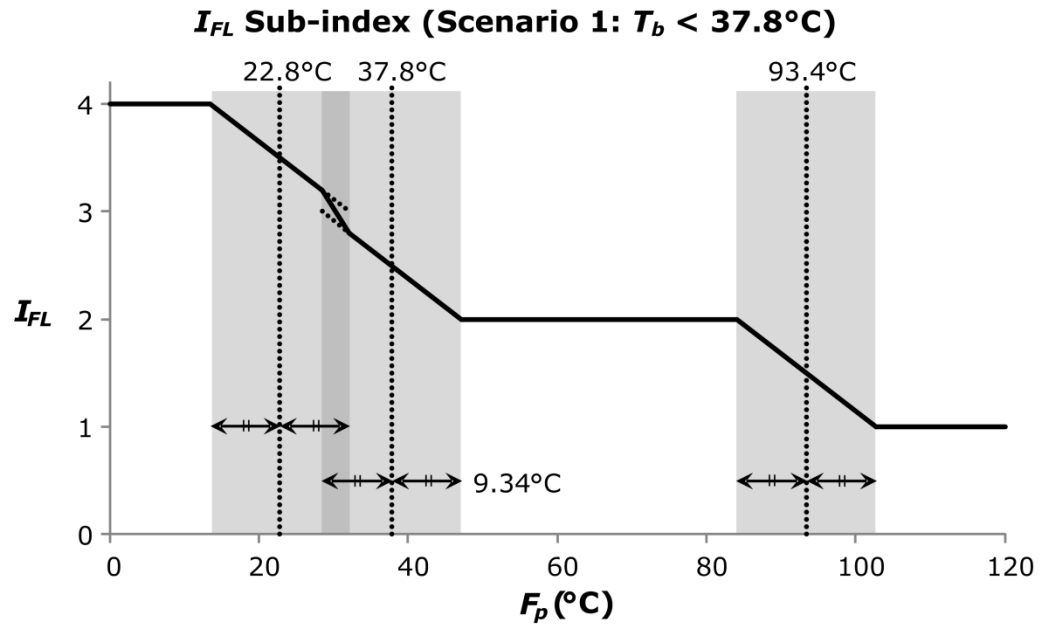


(a)



(b)

**Figure 7.5:** Determining the range to be smoothened in  $I_{FL}$  sub-index



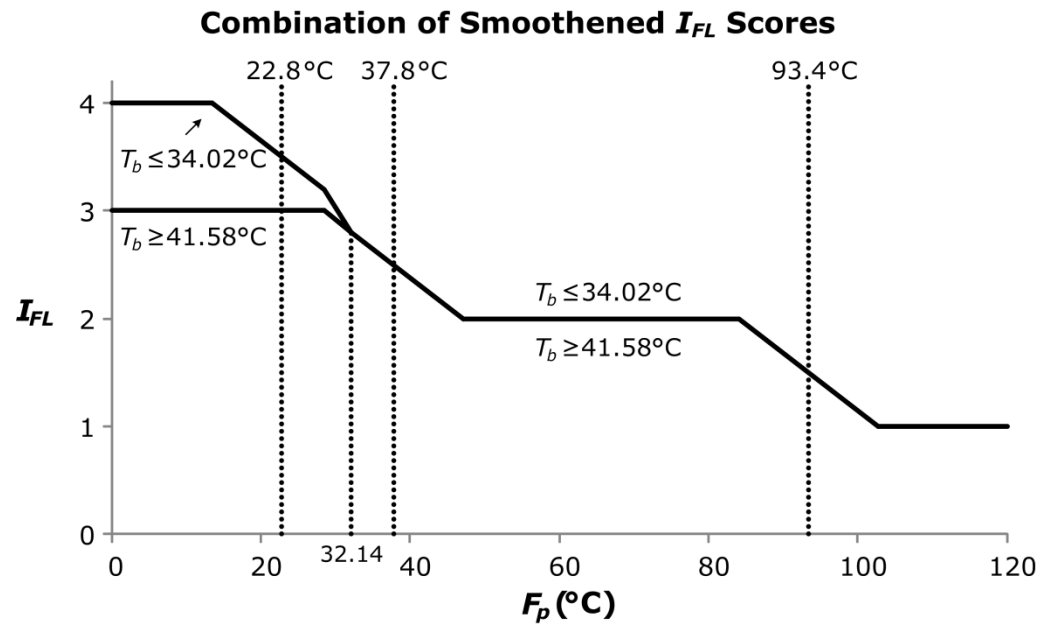
**Figure 7.6:** Smoothened  $I_{FL}$  scores

of the largest margin value ensures that this factor is at least 10% of all  $F_p$  boundaries. Next, a smoothing range is determined for each  $F_p$  boundary value, where the lower bound is calculated by subtracting  $9.34^\circ\text{C}$  from the boundary value, while the upper bound is determined by adding  $9.34^\circ\text{C}$  to

the boundary value. The smoothing regions are shown by the grey zones in Figure 7.5. In these regions, the allocation of scores is modified so that the scores can transit continuously from one value to another. For each region, a linear slope is introduced where it begins from the  $I_{FL}$  score at the lower bound and decreases linearly to the  $I_{FL}$  score at the upper bound. With the application of linear slopes, the smoothened  $I_{FL}$  scores in terms of  $F_p$  are shown in Figure 7.6. However, there exists an overlapping of two smoothing ranges (shown by the darker grey zone) in Figure 7.6(a) as the lower bound for 37.8°C smoothing range is lower than the upper bound for 22.8°C range. As a result, two linear slopes (illustrated by the two dotted lines) are present in the overlapping range. However, any  $F_p$  value in the feasible range can only be represented by a single score. Hence, a composite curve is used on the darker grey region as shown in Figure 7.6(a) to merge the two linear slopes in which the score transits linearly from the higher score at the lower bound (28.46°C) to the lower score at the upper bound (32.14°C). Since the  $I_{FL}$  scores are now smoothened in terms of  $F_p$ , the next step is to further smoothen the scores with respect to  $T_b$ .

As shown by the two scenarios in Figures 7.4 to 7.6,  $T_b$  has one boundary value of 37.8°C. The margin value applied in  $T_b$  is the 10% factor of 37.8°C, which is equivalent to 3.78°C. The smoothened range for  $T_b$  is between 34.02°C and 41.58°C ( $37.8 \pm 3.78^\circ\text{C}$ ). Since Scenario 1 is intended for  $T_b$  value below 37.8°C, while Scenario 2 represents  $I_{FL}$  scores for  $T_b$  value above and equals to 37.8°C; the scores in Figure 7.6(a) now is applied for  $T_b$  value below or equals to 34.02°C, whereas the scores in Figure 7.6(b) is utilised for  $T_b$  value above or equals to 41.58°C. The scores in Figures 7.6(a) and (b) will now serve as the lower and upper bound

values for the  $T_b$  smoothing range respectively. Figure 7.7 shows the combined  $I_{FL}$  scores from Figures 7.6(a) and (b).



**Figure 7.7:** Combined  $I_{FL}$  scores

**Table 7.4:** Smoothened  $I_{FL}$  scores

$F_p$ (°C)	$I_{FL,low}$ ( $T_b \leq 34.02^\circ\text{C}$ )	$I_{FL,up}$ ( $T_b \geq 41.58^\circ\text{C}$ )
$F_p \geq 102.74^\circ\text{C}$	1	1
$84.06^\circ\text{C} \leq F_p \leq 102.74^\circ\text{C}$	$\frac{102.74 - F_p}{18.68} + 1$	$\frac{102.74 - F_p}{18.68} + 1$
$47.14^\circ\text{C} \leq F_p \leq 84.06^\circ\text{C}$	2	2
$32.14^\circ\text{C} \leq F_p \leq 47.14^\circ\text{C}$	$\frac{47.14 - F_p}{18.68} + 2$	$\frac{47.14 - F_p}{18.68} + 2$
$28.46^\circ\text{C} \leq F_p \leq 32.14^\circ\text{C}$	$\frac{0.394(32.14 - F_p)}{3.68} + 2.803$	$\frac{47.14 - F_p}{18.68} + 2$
$13.46^\circ\text{C} \leq F_p \leq 28.46^\circ\text{C}$	$\frac{32.14 - F_p}{18.68} + 3$	3
$F_p \leq 13.46^\circ\text{C}$	4	3

From Figure 7.7, at  $F_p$  below 32.14°C, the  $I_{FL}$  scores at or below the lower bound of  $T_b$  smoothened range (34.02°C) differ from that of the scores at or above the upper bound (41.58°C). The two sets of scores converge at  $F_p$  of 32.14°C, and the scores remain similar above that  $F_p$  value. Therefore, for a molecule with  $F_p$  above or equivalent to 32.14°C, the  $I_{FL}$  score is only dependent on its  $F_p$  regardless of the  $T_b$  value. For a molecule with  $F_p$  lower than 32.14°C, if its  $T_b$  is in the smoothing range (between 34.02°C and 41.58°C), interpolation technique is required to determine its  $I_{FL}$  score. First, both the  $T_b$  lower and upper bound scores are identified from Figure 7.7 based on  $F_p$  value of the molecule. The lower and upper bound scores now represent the  $I_{FL}$  scores at  $T_b$  of 34.02°C and 41.58°C respectively. Based on the  $T_b$  value, its  $I_{FL}$  score can be determined through interpolation between the lower and upper bound scores. For molecules with  $T_b$  below 34.02°C or above 41.58°C, the scores are directly determined from the curves in Figure 7.7. The overall smoothened  $I_{FL}$  scores are summarised in Table 7.4. The other sub-indexes are also smoothened using the same technique, as shown in Tables 7.5 to 7.9.  $I_{MS}$  is the only sub-index that is not smoothened, as it measures the material state of the molecule without the need to divide the properties into sub-ranges. The  $I_{MS}$  score for gas, liquid and solid are 1, 2 and 3 respectively.

Based on Table 7.4, the final  $I_{FL}$  score can be determined by Equation (7.23):

$$I_{FL} = \begin{cases} I_{FL,low} & T_b \leq 34.02^\circ\text{C} \\ \frac{I_{FL,low}(T_b - 34.02) + I_{FL,up}(41.58 - T_b)}{7.56} & 34.02^\circ\text{C} \leq T_b \leq 41.58^\circ\text{C} \\ I_{FL,up} & T_b \geq 41.58^\circ\text{C} \end{cases} \quad (7.23)$$

**Table 7.5:** Smoothened  $I_{EX}$  scores

Parameter	Score Information	Penalty Score
Explosiveness, $I_{EX}$		
$S = (UEL-LEL)$ vol%	$0 \leq S \leq 13$	1
	$13 \leq S \leq 27$	$(S - 13)/14 + 1$
	$27 \leq S \leq 38$	2
	$38 \leq S \leq 52$	$(S - 38)/14 + 2$
	$52 \leq S \leq 63$	3
	$63 \leq S \leq 77$	$(S - 63)/14 + 3$
	$77 \leq S \leq 100$	4

**Table 7.6:** Smoothened  $I_\eta$  scores

Parameter	Score Information	Penalty Score
Viscosity, $I_\eta$	$-1 \leq \log \eta \leq -0.1$	1
	$-0.1 \leq \log \eta \leq 0.1$	$(\log \eta + 0.1)/0.2 + 1$
	$0.1 \leq \log \eta \leq 0.9$	2
	$0.9 \leq \log \eta \leq 1.1$	$(\log \eta - 0.9)/0.2 + 2$
	$1.1 \leq \log \eta \leq 2$	3

**Table 7.7:** Smoothened  $I_V$  scores

Parameter	Score Information	Penalty
Volatility, $I_V$	$T_b \geq 165^\circ\text{C}$	0
	$165^\circ\text{C} \geq T_b \geq 135^\circ\text{C}$	$(165 - T_b)/30$
	$135^\circ\text{C} \geq T_b \geq 65^\circ\text{C}$	1
	$65^\circ\text{C} \geq T_b \geq 35^\circ\text{C}$	$(65 - T_b)/30 + 1$
	$35^\circ\text{C} \geq T_b \geq 15^\circ\text{C}$	2
	$15^\circ\text{C} \geq T_b \geq -15^\circ\text{C}$	$(15 - T_b)/30 + 2$
	$T_b \leq -15^\circ\text{C}$	3

**Table 7.8:** Smoothened  $I_{EL}$  scores

Parameter	Score Information	Penalty Score
Exposure limit, $I_{EL}$	<i>Vapour (ppm)</i>	
	$\log PEL \geq 3.3$	0
	$2.7 \leq \log PEL \leq 3.3$	$(3.3 - \log PEL)/0.6$
	$2.3 \leq \log PEL \leq 2.7$	1
	$1.7 \leq \log PEL \leq 2.3$	$(2.3 - \log PEL)/0.6 + 1$
	$1.3 \leq \log PEL \leq 1.7$	2
	$0.7 \leq \log PEL \leq 1.3$	$(1.3 - \log PEL)/0.6 + 2$
	$0.3 \leq \log PEL \leq 0.7$	3
	$-0.3 \leq \log PEL \leq 0.3$	$(0.3 - \log PEL)/0.6 + 3$
	$\log PEL \leq -0.3$	4

**Table 7.9:** Smoothened  $I_{AH}$  scores

Parameter	Score Information	Penalty Score
Acute health	<i>Oral rat <math>LD_{50}</math> (mg/kg)</i>	
hazard, $I_{AH}$	$D' \geq 3.6311$	0
Let $D' = \log LD_{50}$	$3.0291 \leq D' \leq 3.6311$	$1.5147(3.6311 - D')$
	$2.9709 \leq D' \leq 3.0291$	$3.0294(3.0291 - D')$ $+ 0.9119$
	$2.3689 \leq D' \leq 2.9709$	$1.5147(3.0291 - D') + 1$
	$2.0291 \leq D' \leq 2.3689$	2
	$1.3689 \leq D' \leq 2.0291$	$1.5147(2.0291 - D') + 2$
	$1.0291 \leq D' \leq 1.3689$	3
	$0.3689 \leq D' \leq 1.0291$	$1.5147(1.0291 - D') + 3$
	$D' \leq 0.3689$	4



## 7.6.4 Molecular Design

In this case study, the selection of first-order molecular building blocks is similar to those chosen in Section 6.4.4. To ensure that feasible molecules are synthesised, structural constraints listed in Sections 5.3.4.1 and 6.4.4 are implemented in the model.

## 7.6.5 Optimisation Model Formulation

In the final stage, the four design objective properties are then converted into their respective property operator. The purpose of altering the form is to reduce the non-linearity equations in the model. The next step is to optimise the property operators of the four objective properties, which are  $\Omega_{Tb}$ ,  $\Omega_{Hv}$ ,  $\Omega_{Ra}$ , and  $\Omega_{SHI,w}$ . Four degrees of satisfaction,  $\lambda_{Tb}$ ,  $\lambda_{Hv}$ ,  $\lambda_{Ra}$ , and  $\lambda_I$  are introduced to  $T_b$ ,  $H_v$ ,  $R_a$ , and  $I_{SHI,w}$  respectively for each of their linear membership function as shown by Equations (7.24) to (7.27). The objective function of the model is given by Equation (5.20), subjected to constraints (4.8) and (5.21). For multiple molecular solutions, integer cuts method is conducted whereby additional constraints are added into the model to synthesise alternate molecular structures.

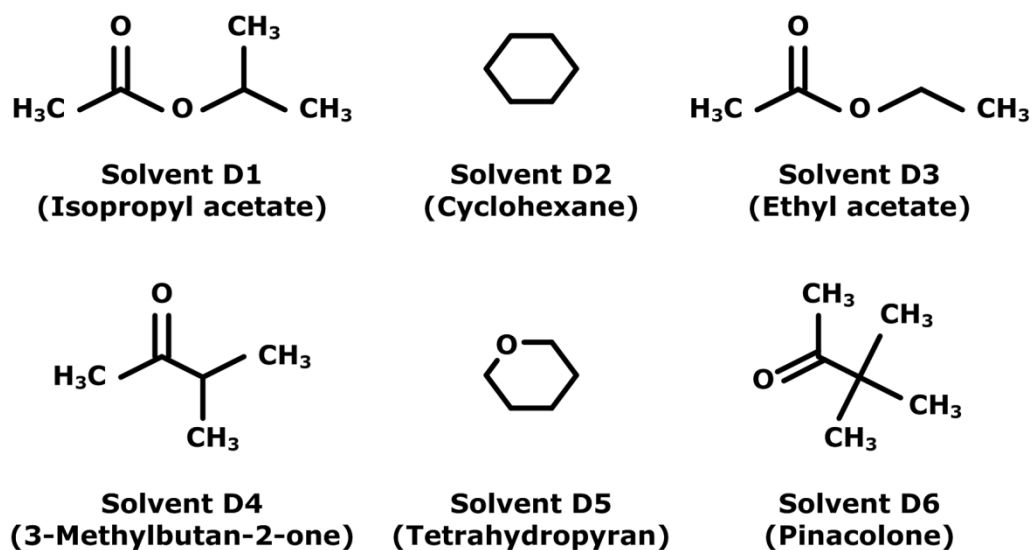
$$\frac{7.3889 - \Omega_{Tb}}{7.3889 - 3.0238} = \lambda_{Tb} \quad (7.24)$$

$$\frac{36.88 - \Omega_{Hv}}{36.88 - 11.6891} = \lambda_{Hv} \quad (7.25)$$

$$\frac{15.8991 - R_a}{15.8991 - 4.0542} = \lambda_{Ra} \quad (7.26)$$

$$\frac{2.6553 - I_{SHI,w}}{2.6553 - 1.7463} = \lambda_I \quad (7.27)$$

## 7.6.6 Results and Discussions



**Figure 7.8:** The generated solvents with their molecular structures

The optimisation model is solved using LINGO 14.0 with a computational time of 86 minutes for the first generated solution. The top six solvents generated by the proposed optimisation model are shown in Figure 7.8. The solvents can be categorised into esters (D1 and D3), ketones (D4 and D6) and monocyclic compounds (D2 and D5). Their estimated properties are shown in Table 7.10, while the sub-index values are summarised in Table 7.11. Solvent D1 is the best solution as it has low  $I_{SHI,w}$  and moderate  $T_b$  and  $R_a$ . When comparing the four objective properties individually, solvent D2 performs the best in terms of lowest  $H_v$  and  $R_a$ ; while solvent D3 offers the lowest  $I_{SHI,w}$  and  $T_b$ . Hence, solvent D2 can be regarded as the best candidate substance in the case where the decision maker priorities property functionalities and performance, while another decision maker can opt for solvent D3 which offers the best safety and health features. In their work, Yara-Varón et al. (2016a) reported that ethyl acetate (solvent D3 in this case study) offers similar carotenoid extraction yield as compared to that of hexane. Therefore, solvents D1, D2

and D5 with lower  $R_a$  than solvent D3 should exhibit better carotenoid extraction than hexane. Even though solvent D6 does not perform the best with regards to the four objectives, it still offers the highest  $F_p$ .

**Table 7.10:** The six generated solvents with their properties

Solvent	$\lambda$	$I_{SHI,w}$	$R_a$	$H_v$ (kJ/mol)	$T_b$ (°C)	$F_p$ (°C)	$LC_{50}$ (mg/l)
D1	0.697	1.957	7.432	34.32	86.2	9.6	114.6
D2	0.687	2.031	4.147	29.08	80.1	-11.9	176.5
D3	0.687	1.946	7.761	32.89	68.7	4.5	208.0
D4	0.617	2.024	8.593	30.88	88.6	6.9	853.1
D5	0.612	2.099	6.804	31.56	89.3	-2.6	875.7
D6	0.597	2.049	8.208	32.07	109.5	14.0	521.9

**Table 7.11:** The six generated solvents with their sub-index scores

Solvent	$I_{FL}$	$I_{EX}$	$I_{\eta}$	$I_{MS}$	$I_V$	$I_{EL}$	$I_{AH}$	$I_{SHI}$	$I_{SHI,w}$
D1	3	1	1	2	1	1.09	0.81	9.90	1.957
D2	3	1	1	2	1	0.86	1.54	10.41	2.031
D3	3	1	1	2	1	1.02	0.82	9.83	1.946
D4	3	1	1	2	1	1.34	1.21	10.54	2.024
D5	3	1	1	2	1	1.58	1.56	11.14	2.099
D6	3	1	1	2	1	1.48	1.23	10.71	2.049

In terms of their individual sub-index scores, the difference between their safety and health performance is mainly contributed by  $I_{EL}$  and  $I_{AH}$ . These two sub-indexes measured the toxicity of chemicals, where  $I_{EL}$  examines the chronic toxicity through inhalation, while  $I_{AH}$  assesses acute toxicity through ingestion.  $I_{SHI}$  in Table 7.11 is calculated by summing up the seven sub-index values of each solvent. Among the seven sub-index

values in Table 7.11, the one with the largest penalty is  $I_{FL}$ , which indicates that the solvents are easily flammable. In the event where a molecule with lower flammability is required, additional integer cuts are introduced to the optimisation model. The best solvent with a lower  $I_{FL}$  score (less than or equals to 2) is 1-ethoxy-2-butanone with a  $\lambda$  value of 0.463, where its  $I_{SHI,w}$ ,  $R_a$ ,  $H_v$ ,  $T_b$ , and  $F_p$  are equivalent to 2.144, 6.599, 37.15 kJ/mol, 137.7°C, and 53.0°C respectively. Its  $R_a$  is lower than the top six solvents in Table 7.10, but it has relatively high  $T_b$  and  $H_v$ , which means that more energy is needed to vaporise the solvent to recover carotenoids. Even though its  $I_{FL}$  has an improved value of 2, but its  $I_{EL}$  score has worsened to 3, indicating that it has higher toxicity than the top six solvents. This also increases its overall inherent hazard level as demonstrated by its high  $I_{SHI,w}$ .

The introduction of index smoothing and prioritisation in this chapter enhances the representation of molecular safety and health performance through the calculation of the adjusted index score,  $I_{SHI,w}$ . The improvement in safety and health measurement can be demonstrated by comparing the results in Table 7.11 with the calculation of  $I_{SHI,i}$  of the six solvents by omitting both index smoothing and prioritisation. In other words, the initial index scoring schemes are used to calculate  $I_{SHI,i}$  by summing up the sub-index scores without allocating weight factors to them. The calculated  $I_{SHI,i}$  and the individual sub-index values are shown in Table 7.12. In Table 7.12, the  $I_{FL}$ ,  $I_{EX}$ ,  $I_{\eta}$ ,  $I_{MS}$  and  $I_V$  scores of the six solvents remain unchanged, as their property values do not fall near to the property boundaries of the five sub-indexes. As for  $I_{EL}$  and  $I_{AH}$ , the results in Table 7.11 show that the property values ( $PEL$  and  $LD_{50}$ ) of the solvents fall near to the property boundaries or the edges of their respective property sub-ranges, so their  $I_{EL}$  and  $I_{AH}$  scores are not discrete values. In

Table 7.12, all solvents but A5 have a similar  $I_{EL}$  score of one, while the  $I_{AH}$  scores of all solvents but A2 and A5 share the same value of one. Solvents A1, A3, A4 and A6 exhibit the lowest  $I_{SHI,i}$  score of 10, while A2 and A5 possess higher hazard with total index scores of 11 and 12 respectively. As four solvents share the same  $I_{SHI,i}$ , it is difficult to distinguish their inherent hazard level. This limitation is addressed with the proposed index smoothing and prioritisation in this paper, where the hazard level of each solvent can be calculated by  $I_{SHI,w}$  shown in Table 7.12.

**Table 7.12:** The six generated solvents (without index smoothing and prioritisation) with their sub-index scores

Solvent	$I_{FL}$	$I_{EX}$	$I_{\eta}$	$I_{MS}$	$I_V$	$I_{EL}$	$I_{AH}$	$I_{SHI,i}$
D1	3	1	1	2	1	1	1	10
D2	3	1	1	2	1	1	2	11
D3	3	1	1	2	1	1	1	10
D4	3	1	1	2	1	1	1	10
D5	3	1	1	2	1	2	2	12
D6	3	1	1	2	1	1	1	10

From the results, the smoothing of scores at property boundary offers a better comparison of hazard level among the solvents. For instance, the calculated  $I_{SHI,i}$  values in Table 7.12 show that solvent A6 is less hazardous than A2. However, the calculated  $I_{SHI,w}$  values in Table 7.11 demonstrate otherwise. This is due to the fact the  $PEL$  and  $LD_{50}$  (used to measure  $I_{EL}$  and  $I_{AH}$  respectively) of solvent A6 are near to the upper boundary (higher degree of hazard) of their respective sub-ranges, thus both its  $I_{EL}$  and  $I_{AH}$  scores as shown in Table 7.11 are penalised more than those in Table 7.12. The same happens to solvent A2 where its  $I_{AH}$  score in Table 7.11 is penalised more than that in Table 7.12. Meanwhile, the  $PEL$  of

solvent A2 falls near to the lower boundary (lower degree of hazard) of its  $I_{EL}$  sub-range, thus its  $I_{EL}$  score as displayed in Table 7.11 is penalised less than that in Table 7.12. This results in the total  $I_{SHI,w}$  of solvent A6 to be greater than that of A2. The smoothing of scores also addresses the limitation of comparing two solvents with property values near to one another but are separated by a sharp property boundary. For instance, the  $I_{EL}$  scores of solvents A5 and A6 without index smoothing are two and one respectively, in which their score difference is one. Their  $PEL$  values are close to one another but they are both separated by a shape scoring edge, which results in their  $I_{EL}$  scores to be different. After the smoothing of sub-indexes, their  $I_{EL}$  score difference is reduced to 0.1, as determined from Table 7.11. Hence, this revision of the sub-index scorings allows the comparison of molecular hazard among different alternatives to be done more accurately.

Meanwhile, the introduction of index prioritisation is performed by assigning different weights to sub-indexes depending on their severity rank. As one of the design objectives is to minimise  $I_{SHI,w}$ , the optimisation model also takes into account the severity of the individual sub-index score when quantifying the overall molecular hazard. Since higher-scoring sub-index scores are assigned with larger weights, this forces the model to minimise the severity of each sub-index score. In the case where weight factors are not introduced to the sub-indexes, the model will only minimise the summation of sub-index scores ( $I_{SHI}$ ). As long as  $I_{SHI}$  is minimised, it does not take into account the severity of each sub-index score. A molecule with minimised  $I_{SHI}$  can still carry a severe-scoring sub-index that is being compensated by few other low-scoring sub-indexes, which is able to lower its  $I_{SHI}$ . To illustrate the improvement brought about by index prioritisation, the values of  $I_{SHI}$  are compared with their corresponding  $I_{SHI,w}$  values in

Table 7.11. It is noted that a higher  $I_{SHI}$  value does not guarantee a larger  $I_{SHI,w}$  value. Even though solvent D2 has lower  $I_{SHI}$  than that of solvent D4, it does not display the same relationship for their corresponding  $I_{SHI,w}$  values. For solvent D2, its  $I_{AH}$  value is higher than its  $I_{EL}$  score, while it shows the opposite for solvent D4. Since solvent D2 has a more severe  $I_{AH}$  score (1.54) than the  $I_{EL}$  score of solvent D4 (1.34), solvent D2 is penalised more with the application of Equation (7.7) to determine  $I_{SHI,w}$ . Thus, the results prove that the adjusted index score takes into account the severity of each sub-index to help distinguish the safety and health attributes of the molecules.

In this chapter, the weight factors introduced to the sub-indexes are determined through the application of pairwise comparison matrix as shown in Table 7.3. However, the allocation of the numerical value,  $a_{ij}$  for each pairwise comparison in the matrix is rather subjective, as different expert may assign different values of  $a_{ij}$  depending on his or her judgement. In this case, an alternate matrix is employed to generate a new set of weight factors for molecular hazard quantification to analyse the sensitivity of the results generated by the optimisation model. Higher importance is given to the top two sub-indexes with the two largest scorings. The new weight factors are 0.4357, 0.2413, 0.1229, 0.0823, 0.0547, 0.0369 and 0.0261 (representing  $w_1$  to  $w_7$  respectively). The two largest weights have increased, while the remaining weights have been reduced. Using the new set of weights to calculate  $I_{SHI,w}$ , the top six solvents with their properties are shown in Table 7.13.

According to Table 7.13, the top five solvents are similar to five of the six solvents generated using the initial set of weights (as shown in Figure 7.8 and Table 7.10). The ranking of the five aforementioned

solvents differ slightly as the calculated  $I_{SHI,w}$  of the solvents in Table 7.13 are different from those in Table 7.11. The new molecule in the top six list is solvent D7 with an IUPAC name of 2-pentanone, which replaces solvent D5 in Table 7.11 as the latter has a larger  $I_{SHI,w}$  value (2.230) calculated from new set of weights. It is observed that under the results generated using the new weight factors, solvent exhibiting better safety and health performance (lower  $I_{SHI,w}$ ) is ranked higher compared to the results obtained with the initial weight factors. Hence, the variation in pairwise comparison matrix used to determine weights for molecular hazard quantification has an impact on the results obtained from the optimisation model. In order to produce a pairwise comparison matrix which minimises bias in decision-making process and better reflects the actual significance impact of each sub-index, it is preferable to acquire the expertise judgements from different individuals to compute the geometric mean for the representation of overall comparison judgements.

**Table 7.13:** The six generated solvents with their properties (using alternate pairwise comparison matrix)

Solvent	$\lambda$	$I_{SHI,w}$	$R_a$	$H_v$ (kJ/mol)	$T_b$ (°C)
D1	0.695	2.119	7.432	34.32	86.2
D3	0.687	2.110	7.761	32.89	68.7
D2	0.641	2.176	4.147	29.08	80.1
D4	0.617	2.171	8.593	30.88	88.6
D6	0.597	2.191	8.208	32.07	109.5
D7	0.597	2.175	8.832	31.78	103.8



## 7.7 Summary

In this chapter, inherent safety and health sub-indexes are integrated into the CAMD framework to measure the safety and health performance of the generated molecules. The main highlight is the development of enhanced method to quantify the overall safety and health attributes of the molecules by applying OWA operator to conditionally assign weight factors to the sub-indexes based on severity rank. Higher severity sub-index score is given a higher weight to further penalise the overall risk level of the molecules. AHP approach is employed to assist in determining the weight factors given to sub-indexes with different severity levels. Besides, the discontinuity of score at the boundary value is managed by smoothing the allocation of scores at the boundary value region. The smoothening region is determined by using a safety/health margin of 10%, and linear transition slope is introduced for the scores allocation. A case study has been considered to replace hexane for the extraction of carotenoids found in PPF. The results show that the top six generated solvents offer high performance with respect to product functionality and the safety and health aspects. The adjusted index score,  $I_{SHI,w}$  as calculated by the OWA operator takes into account the severity of all sub-index scores to enhance the representation of overall hazard level demonstrated by the solvents.

## **CHAPTER 8**

### **CONCLUSIONS**

This chapter concludes all the research works that have been proposed throughout the PhD study. The research gaps in Section 1.3 have been addressed through the completion of research scopes presented in Section 1.4. The concept of inherent safety and occupational health has been successfully adopted in the CAMD framework to formulate molecules with the reduced risk level. However, several works that are not conducted in this research can be addressed in the future to further enhance the integration of safety and health in the CAMD framework.

#### **8.1 Achievements**

Many past industrial catastrophes have caused by the mishandling of hazardous and reactive substances. One way to prevent the potential recurrence of such disasters is to replace those dangerous materials with less harmful ones. The concept of inherent safety and occupational health can be employed, which strives to reduce the amount of hazardous chemicals used in a process plant. The search for a suitable chemical substitute can be effectively carried out by CAMD, which is a useful tool to explore molecular structures that best fulfil the desired functionalities and characteristics. Previously, many CAMD problems did not emphasise strongly about the integration of safety and health aspects as design objectives. Both aspects are often imposed as design constraints, usually represented by flammability and toxicity, to eliminate molecules that are deemed to be dangerous. It is crucial to conduct a trade-off between the

targeted functionalities and the safety and health characteristics, so that the best optimal solution with regards to both criteria can be generated.

The principal objective of this work is to incorporate the safety and health aspects in a conventional chemical product design framework using CAMD technique. Several research gaps present in the current state of CAMD are identified, which result in the establishment of four research scopes in this work. The first scope has been addressed in Chapter 4. The first stage is to develop a CAMD model to generate optimal molecules with the targeted functional performance. In the problem formulation phase, a list of target properties is defined to ensure that the molecules can achieve the desired property functionalities. In the model development phase, the property prediction models for all target properties must be determined to estimate the property values of interest. In the molecular design phase, the appropriate molecular building blocks to form the molecules are selected depending on the types of chemical required for the particular design problem. Structural constraints on the molecular groups are implemented to prevent the formation of infeasible solutions. In the optimisation model phase, the target properties to be optimised are chosen as the design objectives, while the remaining properties will serve as constraints to be fulfilled. Fuzzy optimisation formulation is applied to simultaneously optimise all objectives to generate a set of optimal molecular structures. In the second stage, performance analysis is conducted on the optimal molecules, where they are evaluated with respect to their safety and health attributes. Inherent safety and occupational health sub-indexes are utilised to carry out the assessment, which include flammability, explosiveness, viscosity, material phase, volatility, exposure limit, and acute health hazard. The physical and chemical properties of the molecules are used to measure these sub-indexes. Each sub-index is given

a penalty score to reflect on the molecular hazard level, and the overall safety and health level of a molecule is calculated by the summation of all sub-index scores. Molecules with better scores are selected as the chemical candidates for a specific chemical process application. The outcome of this scope is to evaluate the optimal molecules in terms of multiple safety and health indicators, so that molecules with optimised property functionalities that also demonstrate favourable safety and health performance are produced.

The second research scope has been addressed in Chapter 5. The results in Chapter 4 only guarantee that the molecules are optimised with respect to property functionalities, but not the safety and health criteria. In the second scope, both safety and health aspects are included as design objectives in the molecular design problem. A single-stage CAMD framework has been developed to design molecules that are optimised in terms of property functional performance and safety and health aspects. Similar safety and health sub-indexes are used to evaluate the safety and health attributes of the molecules. As the CAMD programming model is searching for the optimal molecular structure, the corresponding index score must also be simultaneously calculated. Once the model has estimated the molecular properties through property prediction models, disjunctive programming algorithm is applied to convert the input property values to their respective scores. The first and second scopes considered the same case study on solvent design for a gas sweetening process to replace MEA as CO<sub>2</sub>-absorbent. The main limitation of using MEA is its high volatility that results in high make-up solvent cost. One of the design objectives is to address this limitation and the results show that the generated solvents have relatively lower volatility than MEA. The solvents also demonstrate similar or improved safety and health level than MEA.

In Chapter 6, the third research scope is presented where uncertainty management is integrated into the sub-indexes to improve the accuracy of the scorings. As the sub-indexes are measured by the physicochemical properties of the molecules, the properties are estimated through property prediction models. As a result, the sub-index scores given to a molecule is highly dependent on the accuracy of the property prediction models. The deviation of predicted property value from its actual experimental value can result in its corresponding sub-index score to be shifted into an inaccurate value. The issue is most notable when the estimated value shifts to another property sub-range, thus resulting in a different score being assigned to the molecule. The capability and reliability of most property prediction models to estimate accurate property values can be measured in terms of statistical performance indicators. These indicators are applied to manage the property estimation uncertainty on the sub-index scores. Uncertain range is identified at each property boundary, so that the score can change smoothly from the lower bound to the upper bound of the uncertain range. The improved score is able to better reflect the overall hazard level of each molecule under uncertainty.

In the first three scopes, the quantification of the overall molecular safety and health level is determined by the summation of all sub-index values. However, the results from these three scopes show that certain sub-indexes have larger contribution to the final total index score, while others have less impact. In Chapter 7, the fourth scope is presented to improve the current quantification of hazard level. AHP method has been adopted to determine the appropriate weight factors to the sub-indexes, while OWA operator ranks the sub-index values in an ordered manner so that worse performing sub-index with larger score is given a higher weight, and vice versa. This new approach of calculating the overall risk level helps

to generate a more conservative solution and minimises the possibility of forming molecular structures with highly severe scores. Besides, the discontinuity issue in the current sub-index scoring scheme is addressed by smoothing the scores at the property boundaries. This ensures a continuous change of score at any feasible property value to allow effective comparison on safety and health level among alternative solutions. The third and fourth scopes have considered the case study on solvent design for carotenoid extraction. The aim is to replace the commonly used *n*-hexane as the extraction solvent, since it is highly flammable and toxic to aquatic life. From the results, the generated solvents possess compatible characteristics as to those of *n*-hexane, such as high carotenoid solubility and low heat of consumption needed for solvent recovery. One of the proposed solvent, ethyl acetate exhibits the best safety and health level. This ester is also selected as one of the best green solvents carried in other published work for carotenoid extraction.

In conclusion, the proposed methodologies in this research work consider the trade-off between the two principal objectives: property functionalities and safety and health performance. The presented CAMD framework is able to synthesise promising chemicals which not only achieve the desired functionalities, but also demonstrate low safety and health hazard level. The application of chemicals with minimised hazards in process plant can assist in reducing the magnitude of consequences or the likelihood of occurrence of a possible industrial accident.

## **8.2 Future Work**

Throughout this research, there are several important aspects which have been encountered that can be addressed in future works. First,

property prediction uncertainty is currently only managed on the aspects of safety and health. However, this issue of uncertainty should also be addressed on the property functionalities, which may affect the subsequent product-process design. As each estimated property carries a certain extent of uncertainty, the combination of uncertainties may have adverse impact on the quality of design. Besides, the uncertainties on the safety and health indicators may also result in under or over design of certain process equipment. These design situations can either underestimate the actual risk impact of a chemical substance, or overestimate the design of equipment which brings about additional capital costs. Hence, future work can be carried for simultaneous product-process design, and the uncertainties resulted from property prediction can be quantified and analysed to determine the optimal molecular structure with optimum operating process conditions for a specific chemical process application.

Besides, the optimal molecules generated from the CAMD framework should function and behave in the desired manner as their properties satisfy the target performance. However, the design of molecules purely based on theoretical chemical knowledge and hypothesis does not guarantee that these molecules can attain the desired functionality performance when they are practically applied in actual operating process. Thus, an extension of this research work can be considered. The first phase involves the generation of a list of promising molecular structures from CAMD programming. In the next phase, experimental works are conducted to identify the actual physicochemical properties displayed by the shortlisted molecules. The molecules are then screened based on their actual property functionality performance. Subsequently, the final shortlisted molecules are tested with their process performance by carrying out the specific process in a lab-scale basis. The

optimum process conditions can be identified through modelling approach and the actual results can be analysed to compare the performance demonstrated by the conventionally used chemicals and the novel chemical candidates proposed in the work.

Meanwhile, some safety and health parameters are not chosen as sub-indexes in this research work as their scores are assigned based on non-numerical descriptions, which cannot be easily integrated into the mathematical-based CAMD model. Hence, extensive research work has to be done for the inclusion of such parameters to cover a broader range of safety and health aspects, i.e. heat of reaction, corrosiveness, reactivity, and etc. heat This allows the comparison of many safety and health characteristics and it provides the user a greater flexibility to prioritise certain parameters that are deemed to be significantly more important than others, in which such decision depends on the nature of the molecular design problem. Environment-related parameters can also be taken into consideration in future to provide a complete safety, health and environment evaluation on the molecules. This ensures that the designed molecules with reduced risk to the people and community will not result in much catastrophic impact to the environment and its inhabitants.

The economic aspects on the application of the generated molecules in processing plant that includes the safety and occupational health costs should also be taken into consideration in future work. Instead of only quantifying the overall hazard of molecules, the sub-index scores should also be able to provide process designers with some information on the safety and health cost estimations through the installation of protective barriers to control hazards posed by the molecules. Besides, the safety and health level of different process routes used to manufacture each



generated molecule and the cost of manufacturing can also serve as other decision-making criteria. The costs can then be applied to calculate the economic potential for applying a particular molecule in process plant. This economic potential can serve as another design criterion, so that the molecular design problem can cover the aspects of property functionalities, safety, health, and profit margin.

## REFERENCES

- Abedin, F., Roughton, B., Ye, Q., Spencer, P., Camarda, K., 2017. Computer-aided molecular design of water compatible visible light photosensitizers for dental adhesive. *Chemical Engineering Science* 159, 131-139.
- Abidin, M.Z., Rusli, R., Buang, A., Shariff, A.M., Khan, F.I., 2016. Resolving inherent safety conflict using quantitative and qualitative technique. *Journal of Loss Prevention in the Process Industries* 44, 95-111.
- Achenie, L.E.K., Gani, R., Venkatasubramanian, V., 2003. Computer aided molecular design: Theory and practice, first ed. Elsevier, Amsterdam.
- Ahmad, S.I., Hashim, H., Hassim, M.H., 2014. Numerical Descriptive Inherent Safety Technique (NuDIST) for inherent safety assessment in petrochemical industry. *Process Safety and Environmental Protection* 92, 379-389.
- Ahmad, S.I., Hashim, H., Hassim, M.H., 2016. A graphical method for assessing inherent safety during research and development phase of process design. *Journal of Loss Prevention in the Process Industries* 42, 59-69.
- Aissou, M., Chemat-Djenni, Z., Yara-Varón, E., Fabiano-Tixier, A.-S., Chemat, F., 2017. Limonene as an agro-chemical building block for the synthesis and extraction of bioactive compounds. *Comptes Rendus Chimie* 20, 346-358.

- Akash, B.A., Mamlook, R., Mohsen, M.S., 1999. Multi-criteria selection of electric power plants using analytical hierarchy process. *Electric Power Systems Research* 52, 29-35.
- Amid, A., Ghodsypour, S.H., O'Brien, C., 2011. A weighted max-min model for fuzzy multi-objective supplier selection in a supply chain. *International Journal of Production Economics* 131, 139-145.
- Aminbakhsh, S., Gunduz, M., Sonmez, R., 2013. Safety risk assessment using analytic hierarchy process (AHP) during planning and budgeting of construction projects. *Journal of Safety Research*, 46, 99-105.
- Anderson, G.E., 2011. Solvent Extraction. <http://lipidlibrary.aocs.org/OilsFats/content.cfm?ItemNumber=40337> (accessed 06.01.2017).
- Arbel, A., 1987. Venturing into new technological markets. *Mathematical Modelling* 9, 299-308.
- Austin, N.D., Sahinidis, N.V., Trahan, D.W., 2016. Computer-aided molecular design: An introduction and review of tools, applications, and solution techniques. *Chemical Engineering Research and Design* 116, 2-26.
- Behroozsarand, A., Zamaniyan, A., 2011. Multiobjective optimization scheme for industrial synthesis gas sweetening plant in GTL process. *Journal of Natural Gas Chemistry* 20, 99-109.
- Bellman, R.E., Zadeh, L.A., 1970. Decision-making in a fuzzy environment. *Management Sciences* 17, B141-B164.

- Brignole, E.A., Bottini, S., Gani, R., 1986. A strategy for the design and selection of solvents for separation processes. *Fluid Phase Equilibria* 29, 125-132.
- Broughton, E., 2005. The Bhopal disaster and its aftermath: A review. *Environmental Health: A Global Access Science Source* 4.
- Buxton, A., Livingston, A.G., Pistikopoulos, E.N., 1999. Optimal design of solvent blends for environmental impact minimization. *AIChE Journal* 45, 817-843.
- Cave, S.R., Edwards, D.W., 1997. Chemical process route selection based on assessment of inherent environmental hazard. *Computers & Chemical Engineering* 21, S965-S970.
- Chavali, S., Lin, B., Miller, D.C., Camarda, K.V., 2004. Environmentally-benign transition metal catalyst design using optimization techniques. *Computers & Chemical Engineering* 28, 605-611.
- Chemmangattuvalappil N.G., Eden, M.R., 2013. A novel methodology for property-based molecular design using multiple topological indices. *Industrial & Engineering Chemistry Research* 52, 7090-7103.
- Chemmangattuvalappil, N.G., Solvason, C.C., Bommareddy, S., Eden, M.R., 2010. Reverse problem formulation approach to molecular design using property operators based on signature descriptors. *Computers & Chemical Engineering* 34, 2062-2071.

- Chen, H., Pittman, W.C., Hatanaka, L.C., Harding, B.Z., Boussouf, A., Moore, D.A., Milke, J.A., Mannan, M.S., 2015. Integration of process safety engineering and fire protection engineering for better safety performance. *Journal of Loss Prevention in the Process Industries* 37, 74-81.
- Chong, F.K., Andiappan, V., Ng, D.K.S., Foo, D.C.Y., Eljack, F.T., Atilhan, M., Chemmangattuvalappil N.G., 2017. Design of ionic liquid as carbon capture solvent for a bioenergy system: Integration of bioenergy and carbon capture systems. *ACS Sustainable Chemistry & Engineering* 5, 5241-5252.
- Choo, Y.M., Yap, S.C., Ooi, C.K., Ma, A.N., Goh, S.H., Ong, A.S.H., 1996. Recovered oil from palm-pressed fiber: A good source of natural carotenoids, vitamin E, and sterols. *Journal of the American Oil Chemists' Society* 73, 599-602.
- Churi, N., Achenie, L.E., 1996. Novel mathematical programming model for computer aided molecular design. *Industrial & Engineering Chemistry Research* 35, 3788-3794.
- Cignitti, S., Zhang, L., Gani, R., 2015. Computer-aided framework for design of pure, mixed and blended products. *Computer Aided Chemical Engineering* 37, 2093-2098.
- Constantinou, L., Bagherpour, K., Gani, R., Klein, J.A., Wu, D.T., 1996. Computer aided product design: problem formulations, methodology and applications. *Computers & Chemical Engineering* 20, 685-702.

- Constantinou, L., Gani, R., 1994. New group contribution method for estimating properties of pure compounds. *AIChE Journal* 40, 1697-1710.
- Constantinou, L., Gani, R., O'Connell, J.P., 1995. Estimation of the acentric factor and the liquid molar volume at 298 K using a new group contribution method. *Fluid Phase Equilibria* 103, 11-22.
- Conte, E., Gani, R., Ng, K.M., 2011. Design of formulated products: A systematic methodology. *AIChE Journal* 57, 2431-2449.
- Conte, E., Martinho, A., Matos, H.A., Gani, R., 2008. Combined group-contribution and atom connectivity index-based methods for estimation of surface tension and viscosity. *Industrial & Engineering Chemistry Research* 47, 7940-7954.
- Cordella, M., Tugnoli, A., Barontini, F., Spadoni, G., Cozzani, V., 2009. Inherent safety of substances: Identification of accidental scenarios due to decomposition products. *Journal of Loss Prevention in the Process Industries* 22, 455-462.
- Cussler, E.L., Moggridge, G.D., 2001. *Chemical product design*, first ed. Cambridge University Press, Cambridge.
- Dal Prá, V., Soares, J.F., Monego, D.L., Vendruscolo, R.G., Freire, D.M.G., Alexandri, M., Koutinas, A., Wagner, R., Mazutti, M.A., da Rosa, M.B., 2016. Extraction of bioactive compounds from palm (*Elaeis guineensis*) pressed fiber using different compressed fluids. *The Journal of Supercritical Fluids* 112, 51-56.

- Deb, K., 2001. Multi-objective optimization using evolutionary algorithms. John Wiley & Sons, New York.
- Dev, V.A., Chemmangattuvalappil, N.G., Eden, M.R., 2016. Multi-objective computer-aided molecular design of reactants and products. Computer Aided Chemical Engineering 38, 2055-2060.
- Dong, Q., Chirico, R.D., Yan, X., Hong, X., Frenkel, M., 2005. Uncertainty reporting for experimental thermodynamic properties. Journal of Chemical & Engineering Data 50, 546-550.
- Dow Chemical, 1994. Dow's chemical exposure index guide, first ed. American Institute of Chemical Engineers, New York.
- Dubey, D., Mehra, A., 2014. A bipolar approach in fuzzy multi-objective linear programming. Fuzzy Sets and Systems 246, 127-141.
- Duvedi, A.P., Achenie, L.E.K., 1996. Designing environmentally safe refrigerants using mathematical programming. Chemical Engineering Science 51, 3727-3739.
- Edwards, D.W., Lawrence, D., 1993. Assessing the inherent safety of chemical process routes: Is there a relation between plant costs and inherent safety. Chemical Engineering Research and Design 71, 252-258.
- Ee, A.W.L., Shaik, S.M., Khoo, H.H., 2015. Development and application of a Combined Approach for Inherent Safety and Environmental (CAISEN) assessment. Process Safety and Environmental Protection 96, 138-148.

El-Halwagi, M.M., 2012. Sustainable design through process integration: Fundamentals and applications to industrial pollution prevention, resource conservation, and profitability enhancement, first ed. Elsevier, Oxford.

European Palm Oil Alliance, 2016. Palm oil production. <http://www.palmoilandfood.eu/en/palm-oil-production> (accessed 17.02.2017).

Faulon, J.-L., Churchwell, C.J., Visco, D.P., 2003. The signature molecular descriptor. 2. Enumerating molecules from their extended valence sequences. *Journal of Chemical Information and Modeling* 43, 721-734.

Fishburn, P.C., 1967. Additive utilities with incomplete product sets: Application to priorities and assignments. *Operations Research* 15, 537-542.

Folić, M., Adjiman, C.S., Pistikopoulos, E.N., 2008. Computer-aided solvent design for reactions: Maximizing product formation. *Industrial & Engineering Chemistry Research* 47, 5190-5202.

França, L.F., Meireles, M.A.A., 1997. Extraction of oil from pressed palm oil (*Elaes guineensis*) fibers using supercritical CO<sub>2</sub>. *Ciência e Tecnologia de Alimentos* 17, 384.

Fredenslund, A., Jones, R.L., Prausnitz, J.M., 1975. Group-contribution estimates of activity coefficients in nonideal liquid mixtures. *AIChE Journal* 21, 1086-1099.



- Frutiger, J., Marcarie, C., Abildskov, J., Sin, G., 2016a. Group-contribution based property estimation and uncertainty analysis for flammability-related properties. *Journal of Hazardous Materials* 318, 783-793.
- Frutiger, J., Marcarie, C., Abildskov, J., Sin, G., 2016b. A comprehensive methodology for development, parameter estimation, and uncertainty analysis of group contribution based property models—An application to the heat of combustion. *J. Chem. Eng. Data.* 61, 602-613.
- Gani, R., 2004a. Computer-aided methods and tools for chemical product design. *Chemical Engineering Research and Design* 82, 1494-1504.
- Gani, R., 2004b. Chemical product design: challenges and opportunities. *Computers & Chemical Engineering* 28, 2441-2457.
- Gani, R., 2005. Automatic creation of missing groups through connectivity index for pure-component property prediction. *Industrial & Engineering Chemistry Research* 44, 7262-7269.
- Gani, R., Brignole, E.A., 1983. Molecular design of solvents for liquid extraction based on UNIFAC. *Fluid Phase Equilibria* 13, 331-340.
- Gani, R., Nielsen, B., Fredenslund, A., 1991. A group contribution approach to computer-aided molecular design. *AIChE Journal* 37, 1318-1332.
- Gebreslassie, B.G., Diwekar, U.M., 2015. Efficient ant colony optimization for computer aided molecular design: Case study solvent selection problem. *Computers & Chemical Engineering* 78, 1-9.

- Gerbaud, V., Teles Dos Santos M., Pandya, N., Aubry, J.M., 2017. Computer aided framework for designing bio-based commodity molecules with enhanced properties. *Chemical Engineering Science* 159, 177-193.
- Ghosh, D., Chakraborty, D., 2015. A method for capturing the entire fuzzy non-dominated set of a fuzzy multi-criteria optimization problem. *Fuzzy Sets and Systems* 272, 1-29.
- Gnoni, M.G., Bragatto, P.A., 2013. Integrating major accidents hazard into occupational risk assessment: An index approach. *Journal of Loss Prevention in the Process Industries* 26, 751-758.
- Gupta, J.P., Edwards, D.W., 2003. A simple graphical method for measuring inherent safety. *Journal of Hazardous Materials* 104, 15-30.
- Hajipour, S., Satyro, M.A., 2011. Uncertainty analysis applied to thermodynamic models and process design – 1. Pure components. *Fluid Phase Equilibria*, 307, 78-94.
- Handfield, R., Walton, S.V., Sroufe, R., Melnyk, S.A., 2002. Applying environmental criteria to supplier assessment: A study in the application of the Analytical Hierarchy Process. *European Journal of Operational Research* 141, 70-87.
- Hansen, C.M., 2007. Hansen solubility parameters: A user's handbook, second ed. CRC Press, Boca Raton.

- Hansson, S.O., 2010. Promoting inherent safety. *Process Safety and Environmental Protection* 88, 168-172.
- Harper, P.M., Gani, R., 2000. A multi-step and multi-level approach for computer aided molecular design. *Computers & Chemical Engineering* 24, 677-683.
- Harper, P.M., Gani, R., Kolar, P., Ishikawa, T., 1999. Computer-aided molecular design with combined molecular modeling and group contribution. *Fluid Phase Equilibria* 158-160, 337-347.
- Hassim, M.H., Edwards, D.W., 2006. Development of a methodology for assessing inherent occupational health hazards. *Process Safety and Environmental Protection* 84, 378-390.
- Hassim, M.H., Hurme, M., 2010a. Inherent occupational health assessment during process research and development stage. *Journal of Loss Prevention in the Process Industries* 23, 127-138.
- Hassim, M.H., Hurme, M., 2010b. Inherent occupational health assessment during preliminary design stage. *Journal of Loss Prevention in the Process Industries* 23, 476-482.
- Heikkilä, A.-M., 1999. Inherent safety in process plant design. An index based approach. Doctoral Thesis. VTT Publications 384, Technical Research Centre of Finland, Espoo.

- Heintz, J., Belaud, J.-P., Gerbaud, V., 2014a. Chemical enterprise model and decision-making framework for sustainable chemical product design. *Computers in Industry* 65, 505-520.
- Heintz, J., Belaud, J.-P., Pandya, N., Teles Dos Santos, M., Gerbaud, V., 2014b. Computer aided product design tool for sustainable product development. *Computers & Chemical Engineering* 71, 362-376.
- Hukkerikar, A.S., Kalakul, S., Sarup, B., Young, D.M., Sin, G., Gani, R., 2012a. Estimation of environment-related properties of chemicals for design of sustainable processes: Development of group-contribution<sup>+</sup> (GC<sup>+</sup>) property models and uncertainty analysis. *Journal of Chemical Information and Modeling* 52, 2823–2839.
- Hukkerikar, A.S., Sarup, B., Kate, A.T., Abildkov, J., Sin, G., Gani, R., 2012b. Group-contribution<sup>+</sup> (GC<sup>+</sup>) based estimation of properties of pure components: Improved property estimation and uncertainty analysis. *Fluid Phase Equilibria* 321, 25-43.
- Ibrahim, A., Onwuala, A.P., Technologies for extraction of oil from oil-bearing agricultural products: A review. *Journal of Agricultural Engineering and Technology* 13, 58-70.
- Jagtap, H.P., Bewoor, A.K., 2017. Use of Analytic Hierarchy Process methodology for criticality analysis of thermal power plant equipment. *Materials Today: Proceedings* 4, 1927-1936.

- Javadian, N., Maali, Y., Mahdavi-Amiri, N., 2009. Fuzzy linear programming with grades of satisfaction in constraints. *Iranian Journal of Fuzzy Systems* 6, 17-35.
- Joback, K.G., 1989. Designing molecules possessing desired physical property values, Volume 1, Part 1. Doctoral Thesis. Massachusetts Institute of Technology, Massachusetts.
- Joback, K.G., Reid, R.C., 1987. Estimation of pure-component properties from group contributions. *Chemical Engineering Communications* 57, 233-243.
- Johnson, V.S., 2001. Occupational health hazard index for proposed chemical. Master thesis. Loughborough University, Loughborough.
- Kang, J., Zhang, J., Gao, J., 2016. Analysis of the safety barrier function: Accidents caused by the failure of safety barriers and quantitative evaluation of their performance. *Journal of Loss Prevention in the Process Industries* 43, 361-371.
- Karunanithi, A.T., Achenie, L.E.K., Gani, R., 2005. A new decomposition-based computer-aided molecular/mixture design methodology for the design of optimal solvents and solvent mixtures. *Industrial & Engineering Chemistry Research* 44, 4785-4797.
- Karunanithi, A.T., Achenie, L.E.K., Gani, R., 2006. A computer-aided molecular design framework for crystallization solvent design. *Chemical Engineering Science* 61, 1247-1260.

- Kier, L.B., Hall, L.H., 1976. Molecular connectivity VII: Specific treatment of heteroatoms. *Journal of Pharmaceutical Sciences* 65, 1806-1809.
- Khan, F.I., Abbasi, S.A., 1998. Multivariate hazard identification and ranking system. *Process Safety Progress* 17, 157-170.
- Khan, F.I., Amyotte, P.R., 2004. Integrated inherent safety index (I2SI): A tool for inherent safety evaluation. *Process Safety Progress* 23, 136-148.
- Klein, J.A., Wu, D.T., 1992. Computer aided mixture design with specified property constraints. *Computers & Chemical Engineering* 16, S229-S236.
- Kletz, T.A., 1978. What you don't have, can't leak. *Chemistry & Industry*, 278-292.
- Kletz, T.A., 1984. *Cheaper, safer plants, or wealth and safety at work*, The Institution of Chemical Engineers, Rugby.
- Kletz, T.A., 1991. *Plant design for safety: A user-friendly approach*. Hemisphere Publishing Corporation, New York.
- Kluczek, A., 2017. An overall multi-criteria approach to sustainability assessment of manufacturing processes. *Procedia Manufacturing* 8, 136-143.
- Koller, G., Fischer, U., Hungerbühler, K., 2000. Assessing safety, health, and environmental impact early during process development. *Industrial & Engineering Chemistry Research* 39, 960-972.

- Kondo, S., Urano, Y., Tokuhashi, K., Takahashi, A., Tanaka, K., 2001. Prediction of flammability of gases by using *F*-number analysis. *Journal of Hazardous Materials* 82, 113-128.
- Kumar, S., Cho, J.H., Moon, I., 2014. Ionic liquid-amine blends and CO<sub>2</sub>BOLs: Prospective solvents for natural gas sweetening and CO<sub>2</sub> capture technology—A review. *International Journal of Greenhouse Gas Control* 20, 87-116.
- Lampe, M., Stavrou, M., Schilling, J., Sauer, E., Gross, J., Bardow, A., 2015. Computer-aided molecular design in the continuous-molecular targeting framework using group-contribution PC-SAFT. *Computers & Chemical Engineering* 81, 278-287.
- Larsen, A.H., 1986. Data quality for process design. *Fluid Phase Equilibria* 29, 47-58.
- Lau, H.L.N., Choo, Y.M., Ma, A.N., Chuah, C.H., 2008. Selective extraction of palm carotene and vitamin E from fresh palm-pressed mesocarp fiber (*Elaeis guineensis*) using supercritical CO<sub>2</sub>. *Journal of Food Engineering* 84, 289-296.
- Lee, B.I., Kesler, M.G., 1980. Improve vapor pressure prediction. *Hydrocarbon Process* 59, 163-167.
- Leong, C.T., Shariff, A.M., 2009. Process route index (PRI) to assess level of explosiveness for inherent safety quantification. *Journal of Loss Prevention in the Process Industries* 22, 216-221.

- Li, M., Liu, Y., Peng, M., Xie, C., Yang, L., 2016. The decision making method of task arrangement based on analytic hierarchy process for nuclear safety in radiation field. *Progress in Nuclear Energy* 93, 318-326.
- Ma, T., Wang, Q., Larrañaga, M.D., 2013. Correlations for estimating flammability limits of pure fuels and fuel-inert mixtures. *Fire Safety Journal* 56, 9-19.
- Mansfield, D., Hawksley, J., 1998. Improved business performance through inherent safety, health and environmental design and operation: A report of a London conference. *Journal of Cleaner Production* 6, 147-150.
- Maranas, C.D., 1996. Optimal computer-aided molecular design: A polymer design case study. *Industrial & Engineering Chemistry Research* 35, 3403-3414.
- Maranas, C.D., 1997. Optimal molecular design under property prediction uncertainty. *AIChE Journal* 43, 1250-1264.
- Marcoulaki, E.C., Kokossis, A.C., 2000. On the development of novel chemicals using a systematic synthesis approach. Part I. Optimisation framework. *Chemical Engineering Science* 55, 2529-2546.
- Marrero, J., Gani, R., 2001. Group-contribution based estimation of pure component properties. *Fluid Phase Equilibria* 183-184, 183-208.



- Marrero, J., Gani, R., 2002. Group-contribution-based estimation of octanol/water partition coefficient and aqueous solubility. *Industrial & Engineering Chemistry Research* 41, 6623-6633.
- Martin, T.M., Young, D.M., 2001. Prediction of the acute toxicity (96-h LC50) of organic compounds to the fathead minnow (*Pimephales promelas*) using a group contribution method. *Chemical Research in Toxicology* 14, 1378-1385.
- Mathias, P.M., Klotz, H.C., 1994. Take a closer look at thermodynamic property models. *Chemical Engineering Progress* 90, 67-75.
- Mattei, M., Kontogeorgis, G.M., Gani, R., 2013. Design of an emulsion-based personal detergent through a model-based chemical product design methodology. *Computer Aided Chemical Engineering* 32, 817-822.
- Mba, O.I., Dumont, M.-J., Ngadi, M., 2015. Palm oil: Processing, characterization and utilization in the food industry – A review. *Food Bioscience* 10, 26-41.
- Muhammad, A., GadelHak, Y., 2015. Simulation based improvement techniques for acid gases sweetening by chemical absorption: A review. *International Journal of Greenhouse Gas Control* 37, 481-491.
- Nannoolal, Y., Rarey, J., Ramjugernath, D., 2008. Estimation of pure component properties: Part 3. Estimation of the vapor pressure of non-electrolyte organic compounds via group contributions and group interactions. *Fluid Phase Equilibria* 269, 117-133.

National Fire Protection Association, 2007. NFPA 704: Standard system for the identification of the hazards of materials for emergency response. National Fire Protection Association, Quincy.

Neoh, B.K., Thang, Y.M., Zain, M.Z.M., Junaidi, A., 2011. Palm pressed fibre oil: A new opportunity for premium hardstock. *International Food Research Journal* 18, 769-773.

Ng, L.Y., Chemmangattuvalappil, N.G., Ng, D.K.S., 2014a. A multiobjective optimization-based approach for optimal chemical product design. *Industrial & Engineering Chemistry Research* 53, 17429-17444.

Ng, L.Y., Chemmangattuvalappil, N.G., Ng, D.K.S., 2015a. Robust chemical product design via fuzzy optimisation approach. *Computers & Chemical Engineering* 83, 186.

Ng, L.Y., Chong, F.K., Chemmangattuvalappil, N.G., 2015b. Challenges and opportunities in computer-aided molecular design. *Computers & Chemical Engineering* 81, 115-129.

Ng, R.T.L., Ng, D.K.S., Tan, R.R., El-Halwagi, M.M., 2014b. Disjunctive fuzzy optimisation for planning and synthesis of bioenergy-based industrial symbiosis system. *Journal of Environmental Chemical Engineering* 2, 652-664.

Odele, O., Macchietto, S., 1993. Computer aided molecular design: A novel method for optimal solvent selection. *Fluid Phase Equilibria* 82, 47-54.

- Odynocki, B., 1979. Planning the national health insurance policy: An application of the AHP in health policy evaluation and planning. PhD dissertation. University of Pennsylvania, Philadelphia.
- Okoh, P., Haugen, S., 2014. Application of inherent safety to maintenance-related major accident prevention on offshore installations. *Chemical Engineering Transactions* 36, 175-180.
- Ostrovsky, G.M., Achenie, L.E.K., Sinha, M., 2002. On the solution of mixed-integer nonlinear programming models for computer aided molecular design. *Computers & Chemistry* 26, 645-660.
- Ourique, J.E., Silva Telles, A., 1998. Computer-aided molecular design with simulated annealing and molecular graphs. *Computers & Chemical Engineering* 22, S615-S618.
- Palaniappan, C., Srinivasan, R., Tan, R., 2002. Expert system for the design of inherently safer processes. 1. Route selection stage. *Industrial & Engineering Chemistry Research* 41, 6698-6710.
- Palma-Flores, O., Flores-Tlacuahuac, A., Canseco-Melchor, G., 2016. Simultaneous molecular and process design for waste heat recovery. *Energy* 99, 32-47.
- Papadokonstantakis, S., Badr, S., Hungerbühler, K., Papadopoulos, A.I., Damartzis, T., Seferlis, P., Forte, E., Chremos, A., Galindo, A., Jackson, G., Adjiman C.S., 2015. Chapter 11 - Toward sustainable solvent-based postcombustion CO<sub>2</sub> capture: From molecules to conceptual flowsheet design. *Computer Aided Chemical Engineering* 36, 279-310.

- Papadopoulos, A.I., Stijepovic, M., Linke, P., 2010. On the systematic design and selection of optimal working fluids for Organic Rankine Cycles. *Applied Thermal Engineering* 30, 760-769.
- Papadopoulos, A.I., Stijepovic, M., Linke, P., Seferlis, P., Voutetakis, S., 2013. Molecular design of working fluid mixtures for Organic Rankine Cycles. *Computer Aided Chemical Engineering* 32, 289-294.
- Patel, S.J., Ng, D., Mannan, M.S., 2010. Inherently safer design of solvent processes at the conceptual stage: Practical application for substitution. *Journal of Loss Prevention in the Process Industries* 23, 483-491.
- Peters, L., Hussain, A., Follmann, M., Melin, T., Hägg, M.-B., 2011. CO<sub>2</sub> removal from natural gas by employing amine absorption and membrane technology—A technical and economical analysis. *Chemical Engineering Journal* 172, 952-960.
- Pistikopoulos, E.N., Stefanis, S.K., 1998. Optimal solvent design for environmental impact minimization. *Computers & Chemical Engineering*, 22, 717-733.
- Promentilla, M.A.B., Furuichi, T., Ishii, K., Tanikawa, N., 2006. Evaluation of remedial countermeasures using the analytic network process. *Waste Management* 26, 1410-1421.
- Promentilla, M.A.B., Kalaw, M.E., Nguyen, H.T., Aviso, K.B., Tan, R.R., 2017. A fuzzy programming approach to multi-objective optimization for geopolymer product design. *Computer Aided Chemical Engineering* 40, 1015-1020.

- Rahman, M., Heikkila, A.-M., Hurme, M., 2005. Comparison of inherent safety indices in process concept evaluation. *Journal of Loss Prevention in the Process Industries* 18, 327-334.
- Raman, V.S., Maranas, C.D., 1998. Optimization in product design with properties correlated with topological indices. *Computers & Chemical Engineering* 22, 747-763.
- Ramanathan, R., 2002. Successful transfer of environmentally sound technologies for greenhouse gas mitigation: A framework for matching the needs of developing countries. *Ecological Economics* 42, 117-129.
- Randic, M., 1975. Characterization of molecular branching. *Journal of the American Chemical Society* 97, 6609-6615.
- Rathnayaka, S., Khan, F., Amyotte, P., 2014. Risk-based process plant design considering inherent safety. *Safety Science* 70, 438-464.
- Riazi, M.R., Daubert, T.E., 1980. Simple property prediction. *Hydrocarbon Process* 59, 115-116.
- Saaty, T.L., 1977. A scaling method for priorities in hierarchical structures. *Journal of Mathematical Psychology* 15, 234-281.
- Saaty, T.L., 1980. *The Analytic Hierarchy Process*, Mc-Graw-Hill, New York.
- Saaty, T.L., 1994. Highlights and critical points in the theory and application of the Analytic Hierarchy Process. *European Journal of Operational Research* 74, 426-447.

- Saaty, T.L., Gholamnezhad, H., 1982. High-level nuclear waste management: Analysis of options. *Environment and Planning B: Urban Analytics and City Science* 9, 181-196.
- Saaty, T.L., Kearns, K.P., 1985. *Analytical planning: The organization of systems*, first ed. Pergamon Press, Oxford.
- Satyanarayana, K.C., Abildskov, J., Gani, R., 2009. Computer-aided polymer design using group contribution plus property models. *Computers & Chemical Engineering* 33, 1004-1013.
- Scheffczyk, J., Fleitmamm, L., Schwarz, A., Lampe, M., Bardow, A., Leonhard, K., 2017. COSMO-CAMD: A framework for optimization-based computer-aided molecular design using COSMO-RS. *Chemical Engineering Science* 159, 84-92.
- Schniederjans, M.J., Wilson, R.L., 1991. Using the analytic hierarchy process and goal programming for information system project selection. *Information & Management* 20, 333-342.
- Seider, W.D., Seader, J.D., Lewin, D.R., 2004. *Product and process design principles: Synthesis, analysis, and evaluation*, second ed. Wiley, New York.
- Serna, J., Díaz Martinez, E.N., Narváez Rincón, P.C., Camargo, M., Gálvez, D., Orjuela, Y., 2016. Multi-criteria decision analysis for the selection of sustainable chemical process routes during early design stages. *Chemical Engineering Research and Design* 113, 28-49.

- Seth, R., Mackay, D., Muncke, J., 1999. Estimating the organic carbon partition coefficient and its variability for hydrophobic chemicals. *Environmental Science and Technology* 33, 2390-2394.
- Shakerian, F., Kim, K.-H., Szulejko, J.E., Park, J.-W., 2015. A comparative review between amines and ammonia as sorptive media for post-combustion CO<sub>2</sub> capture. *Applied Energy* 148, 10-22.
- Shariff, A.M., Leong, C.T., 2009. Inherent risk assessment—A new concept to evaluate risk in preliminary design stage. *Process Safety and Environmental Protection* 87, 371-376.
- Shariff, A.M., Zaini, D., 2013. Inherent risk assessment methodology in preliminary design stage: A case study for toxic release. *Journal of Loss Prevention in the Process Industries* 26, 605-613.
- Singh, S.K., Yadav, S.P., 2015. Modeling and optimization of multi objective non-linear programming problem in intuitionistic fuzzy environment. *Applied Mathematical Modelling* 39, 4617-4629.
- Sinha, M., Achenie, L.E.K., 2003. Blanket wash solvent blend design using interval analysis. *Industrial & Engineering Chemistry Research* 42, 516-527.
- Sinha, M., Achenie, L.E.K., Ostrovsky, G.M., 1999. Environmentally benign solvent design by global optimization. *Computers & Chemical Engineering* 23, 1381-1394.

- Sinnott, R.K., 2005. Chemical engineering, Volume 6: Chemical engineering design, fourth ed. Elsevier, Oxford.
- Srinivasan, R., Natarajan, S., 2012. Developments in inherent safety: A review of the progress during 2001–2011 and opportunities ahead. *Process Safety and Environmental Protection* 90, 389-403.
- Srinivasan, R., Nhan, N.T., 2008. A statistical approach for evaluating inherent benign-ness of chemical process routes in early design stages. *Process Safety and Environmental Protection* 86, 163-174.
- Stavrou, M., Lampe, M., Bardow, A., Gross, J., 2014. Continuous Molecular targeting–computer-aided molecular design (CoMT–CAMD) for simultaneous process and solvent design for CO<sub>2</sub> capture. *Industrial & Engineering Chemistry Research* 53, 18029-18041.
- Tay, D.H.S., Ng, D.K.S., Sammons, N.E., Eden, M.R., 2011. Fuzzy optimization approach for the synthesis of a sustainable integrated biorefinery. *Industrial & Engineering Chemistry Research* 50, 1652-1665.
- Tyler, B.J., 1985. Using the mond index to measure inherent hazards. *Plant/Operations Progress* 4, 172-175.
- Tyler, B.J., Thomas, A.R., Doran, P., Greig, T.R., 1996. A toxicity hazard index. *Chemical Health and Safety*, 19-25.



- Venkatasubramanian, V., Chan, K., Caruthers, J.M., 1994. Computer-aided molecular design using genetic algorithms. *Computers & Chemical Engineering* 18, 833-844.
- Visco, D.P., Chen, J.J., 2017. Chapter 11: The signature molecular descriptor in molecular design: Past and current applications. *Computer Aided Chemical Engineering* 39, 315-343.
- Visco, D.P., Pophale, R.S., Rintoul, M.D., Faulon, J.-L., 2002. Developing a methodology for an inverse quantitative structure-activity relationship using the signature molecular descriptor. *Journal of Molecular Graphics and Modelling* 20, 429-438.
- Wan, Y.K., Ng, R.T.L., Ng, D.K.S., Aviso, K.B., Tan, R.R., 2016. Fuzzy multi-footprint optimisation (FMFO) for synthesis of a sustainable value chain: Malaysian sago industry. *Journal of Cleaner Production* 128, 62-76.
- Wang, T., Hovland, J., Jens, K.J., 2015. Amine reclaiming technologies in post-combustion carbon dioxide capture. *Journal of Environmental Sciences* 27, 276-289.
- Warnasooriya, S., Gunasekera, M.Y., 2017. Assessing inherent environmental, health and safety hazards in chemical process route selection. *Process Safety and Environmental Protection* 105, 224-236.
- Wenham, D., 2002. Occupational health and safety management course module. Centre for Hazard and Risk Management (CHaRM), Loughborough.

- Whiting, W.B., 1996. Effects of uncertainties in thermodynamic data and models on process calculations. *Journal of Chemical & Engineering Data* 41, 935-941.
- Xu, Z.S., Da, Q.L., 2003. An overview of operators for aggregating information. *International Journal of Intelligent Systems* 18, 953-969.
- Yager, R.R., 1988. On ordered weighted averaging aggregation operators in multicriteria decision making. *IEEE Systems, Man, and Cybernetics Society* 18, 183-190.
- Yamamoto, H., Tochigi, K., 2008. Computer-aided molecular design to select foaming agents using a neural network method. *Industrial & Engineering Chemistry Research* 47, 5152-5156.
- Yan, X., Dong, Q., Hong, X., 2003. Reliability analysis of group-contribution methods in predicting critical temperatures of organic compounds. *Journal of Chemical & Engineering Data* 48, 374-380.
- Yara-Varón, E., Fabiano-Tixier, A.S., Balcells, M., Canela-Garayoa, R., Bily, A., Chemat, F., 2016a. Is it possible to substitute hexane with green solvents for extraction of carotenoids? A theoretical versus experimental solubility study. *RSC Advances* 6, 27750-27759.
- Yara-Varón, E., Selka, A., Fabiano-Tixier, A.S., Balcells, M., Canela-Garayoa, R., Bily, A., Touaibia, M., Chemat, F., 2016b. Solvent from forestry biomass. Pinane a stable terpene derived from pine tree byproducts to substitute n-hexane for the extraction of bioactive compounds. *Green Chemistry* 18, 6596-6608.

Yu, K.D.S., Aviso, K.B., Promentilla, M.A.B., Santos, J.R., Tan, R.R., 2016.

A weighted fuzzy linear programming model in economic input-output analysis: An application to risk management of energy system disruptions. *Environment Systems and Decisions* 36, 183-195.

Zadeh, L.A., 1965. Fuzzy sets. *Information and Control* 8, 338-353.

Zhang, L., Cignitti, S., Gani, R., 2015. Generic mathematical programming formulation and solution for computer-aided molecular design. *Computers & Chemical Engineering* 78, 79-84.

Zimmermann, H.-J., 1978. Fuzzy programming and linear programming with several objective functions. *Fuzzy Sets and Systems* 1, 45-55.

# APPENDICES

This chapter illustrates the coding used to develop the CAMD programming model with LINGO software (version 14.0).

## A.1 Lingo Coding (Chapter 4)

```
model:

!Objective function;
max = lamda;

!Linear membership functions for the objectives;
(3.7313-kow_tot)/(3.7313+2.6284) = lamda_p1;
(tb_tot-5.2373)/(9.3078-5.2373) = lamda_p2;

lamda_p1 >= lamda;
lamda_p2 >= lamda;

lamda >= 0;
lamda <= 1;

!Defining the chemical building blocks
n1 = CH3, n2 = CH2, n3 = CH, n4 = OH, n5 = CH2O, n6 = CH2NH2,
n7 = CH2NH, n8 = CHNH, n9 = CH3N, n10 = CH2N;

!n1 to n10 represent numbers of each chemical block;
@GIN(n1); @GIN(n2); @GIN(n3); @GIN(n4); @GIN(n5); @GIN(n6);
@GIN(n7); @GIN(n8); @GIN(n9); @GIN(n10);

n_tot = n1+n2+n3+n4+n5+n6+n7+n8+n9+n10;

C_number = n1+n2+n3+n5+n6+n7+n8+n9+n10;
H_number = n1*3+n2*2+n3+n4+n5*2+n6*4+n7*3+n8*2+n9*3+n10*2;
N_number = n6+n7+n8+n9+n10;
O_number = n4+n5;

!Group contribution for normal boiling point;
tb1 = 0.8853; @free(tb1);
tb2 = 0.5815; @free(tb2);
tb3 = -0.0039; @free(tb3);
tb4 = 2.1385; @free(tb4);
tb5 = 0.9999; @free(tb5);
tb6 = 2.3212; @free(tb6);
tb7 = 1.3838; @free(tb7);
tb8 = 0.7116; @free(tb8);
tb9 = 1.0505; @free(tb9);
tb10 = 0.4199; @free(tb10);

!Group contribution for enthalpy of vaporization;
hv1 = 2.2643;
hv2 = 4.7607;
hv3 = 5.0336;
```

```

hv4 = 24.1639;
hv5 = 8.5931;
hv6 = 16.3428;
hv7 = 11.9165;
hv8 = 10.9079;
hv9 = 13.6543;
hv10 = 7.4283;

!Group contribution for liquid molar volume;
mv1 = 0.0241;
mv2 = 0.0165;
mv3 = 0.0086;
mv4 = 0.0044;
mv5 = 0.0228;
mv6 = 0.0281;
mv7 = 0.026;
mv8 = 0.0209;
mv9 = 0.0259;
mv10 = 0.0187;

!Group contribution for octanol-water partition coefficient
(Kow);
kow1 = 0.3008; @free(kow1);
kow2 = 0.4352; @free(kow2);
kow3 = 0.3837; @free(kow3);
kow4 = -1.0185; @free(kow4);
kow5 = -0.1449; @free(kow5);
kow6 = -1.465; @free(kow6);
kow7 = -0.9465; @free(kow7);
kow8 = -0.4419; @free(kow8);
kow9 = -0.3519; @free(kow9);
kow10 = -0.6373; @free(kow10);

!Group contribution for melting point (Tm);
tm1 = 0.6699; @free(tm1);
tm2 = 0.2992; @free(tm2);
tm3 = -0.2943; @free(tm3);
tm4 = 3.2702; @free(tm4);
tm5 = 0.7649; @free(tm5);
tm6 = 3.4368; @free(tm6);
tm7 = 2.0673; @free(tm7);
tm8 = 1.6571; @free(tm8);
tm9 = 0.9396; @free(tm9);
tm10 = -0.1982; @free(tm10);

!Group contribution for viscosity;
vis1 = -1.0278; @free(vis1);
vis2 = 0.2125; @free(vis2);
vis3 = 1.318; @free(vis3);
vis4 = 1.3057; @free(vis4);
vis5 = 0.6134; @free(vis5);
vis6 = 0.2902; @free(vis6);
vis7 = 1.0512; @free(vis7);
vis8 = 1.8378; @free(vis8);
vis9 = 0.8715; @free(vis9);
vis10 = 1.4376; @free(vis10);

!Group contribution for molecular weight (MW);
mw1 = 15.035;
mw2 = 14.027;
mw3 = 13.019;

```

```

mw4 = 17.007;
mw5 = 30.026;
mw6 = 30.05;
mw7 = 29.042;
mw8 = 28.034;
mw9 = 29.042;
mw10 = 28.034;

!Property Constraints;
!Equation for normal boiling point;
tb_tot =
n1*tb1+n2*tb2+n3*tb3+n4*tb4+n5*tb5+n6*tb6+n7*tb7+n8*tb8+n9*tb9+
n10*tb10; @free(tb_tot);

!Equation for enthalpy of vaporization;
hv_tot =
n1*hv1+n2*hv2+n3*hv3+n4*hv4+n5*hv5+n6*hv6+n7*hv7+n8*hv8+n9*hv9+
n10*hv10;

!Equation for liquid molar volume;
mv_tot =
n1*mv1+n2*mv2+n3*mv3+n4*mv4+n5*mv5+n6*mv6+n7*mv7+n8*mv8+n9*mv9+
n10*mv10;

!Equation for octanol-water partition coefficient (Kow);
kow_tot =
n1*kow1+n2*kow2+n3*kow3+n4*kow4+n5*kow5+n6*kow6+n7*kow7+n8*kow8
+n9*kow9+n10*kow10; @free(kow_tot);

!Equation for normal melting point (Tm);
tm_tot =
n1*tm1+n2*tm2+n3*tm3+n4*tm4+n5*tm5+n6*tm6+n7*tm7+n8*tm8+n9*tm9+
n10*tm10; @free(tm_tot);

!Equation for viscosity;
vis_tot =
n1*vis1+n2*vis2+n3*vis3+n4*vis4+n5*vis5+n6*vis6+n7*vis7+n8*vis8
+n9*vis9+n10*vis10; @free(vis_tot);
vis_tot <= 4.60517;
vis_tot >= -2.302585;

!Equation for molecular weight;
mw =
n1*mw1+n2*mw2+n3*mw3+n4*mw4+n5*mw5+n6*mw6+n7*mw7+n8*mw8+n9*mw9+
n10*mw10;

!Properties upper and lower boundaries;
tb_tot <= 12.787892;
tb_tot >= 5.22893;
tm_tot <= 7.977927;
tm_tot >= 4.262302;
hv_tot <= 518.3873;
hv_tot >= 40.3873;
mv_tot <= 0.208;
mv_tot >= 0.024;
mw <= 250;
mw >= 60;
vis_tot <= 6.131226;

!Molecular structure constraints;
n1+n2+n3+n4+n5+n6+n7+n8+n9+n10>0; !a molecule must be formed;

```

```

n6+n7+n8+n9+n10=1; !N aton can only appear once;

!Special constraints on OH and CH2O;
n5<=1;
(0-0.5)*I_n5 < n5-0.5;
n5-0.5 <= (1-0.5)*(1-I_n5);

n4<=I_n5*3+(1-I_n5); @BIN(I_n5);

!Free bonds (valence) for each group;
val1 = 1;
val2 = 2;
val3 = 3;
val4 = 1;
val5 = 2;
val6 = 1;
val7 = 2;
val8 = 3;
val9 = 2;
val10 = 3;

!Structural constraint, the molecule generated must not contain
free bonds;
(n1*val1+n2*val2+n3*val3+n4*val4+n5*val5+n6*val6+n7*val7+n8*val
8+n9*val9+n10*val10) - (2*(n1+n2+n3+n4+n5+n6+n7+n8+n9+n10-1)) =
0;

END

```

## A.2 Lingo Coding (Chapter 5)

```

model:

!Objective function;

max = lamda;

!Linear membership functions for objectives;
(13-I_SHI)/(13-10) = lamda_I;
(3.5933-kow_tot)/(3.5933+2.6284) = lamda_p1;
(tb_tot-5.2373)/(9.6504-5.2373) = lamda_p2;

lamda_I >= lamda;
lamda_p1 >= lamda;
lamda_p2 >= lamda;

lamda >= 0;
lamda <= 1;

!Defining the chemical building blocks
1st-order groups
n1 = CH3, n2 = CH2, n3 = CH, n4 = OH, n5 = CH2O, n6 = CH2NH2,
n7 = CH2NH, n8 = CHNH, n9 = CH3N, n10 = CH2N

2nd-order groups
n201 = (CH3)2CH, n202 = CH(CH3)CH(CH3), n203 = CHOH, n204 =
CH(OH)CH(OH), n205 = CH(OH)CH2(OH), n207 = CH(OH)CH2(NH2), n209
= CH(OH)CH2(NH),

```

```

n210 = CH2(OH)CH2(NH), n211 = CH(OH)CH(NH), n212 =
CH2(OH)CH(NH), n213 = CH(OH)CH2(N), n214 = CH2(OH)CH2(N), n215
= CH(OH)CH(N), n216 = CH2(OH)CH(N);

@GIN(n1); @GIN(n2); @GIN(n3); @GIN(n4); @GIN(n5); @GIN(n6);
@GIN(n7); @GIN(n8); @GIN(n9); @GIN(n10);
@GIN(n201); @GIN(n202); @GIN(n203); @GIN(n204); @GIN(n205);
@GIN(n207); @GIN(n209); @GIN(n210); @GIN(n211); @GIN(n212);
@GIN(n213); @GIN(n214);
@GIN(n215); @GIN(n216);

n_tot = n1+n2+n3+n4+n5+n6+n7+n8+n9+n10;
n2_tot =
n201+n202+n203+n204+n205+n207+n209+n210+n211+n212+n213+n214+n21
5+n216;
@BND(1, n_tot, 13);

C_number = n1+n2+n3+n5+n6+n7+n8+n9+n10;
H_number = n1*3+n2*2+n3+n4+n5*2+n6*4+n7*3+n8*2+n9*3+n10*2;
N_number = n6+n7+n8+n9+n10;
O_number = n4+n5;

!Group contribution for normal boiling point;
tb1 = 0.8853; @free(tb1);
tb2 = 0.5815; @free(tb2);
tb3 = -0.0039; @free(tb3);
tb4 = 2.1385; @free(tb4);
tb5 = 0.9999; @free(tb5);
tb6 = 2.3212; @free(tb6);
tb7 = 1.3838; @free(tb7);
tb8 = 0.7116; @free(tb8);
tb9 = 1.0505; @free(tb9);
tb10 = 0.4199; @free(tb10);
tb201 = 0.0071; @free(tb201);
tb202 = 0.1667; @free(tb202);
tb203 = -0.1193; @free(tb203);
tb204_05 = 0.1944; @free(tb204_05);
tb207_14 = 0.3136; @free(tb207_14);

!Group contribution for enthalpy of vaporization;
hv1 = 2.2643; @free(hv1);
hv2 = 4.7607; @free(hv2);
hv3 = 5.0336; @free(hv3);
hv4 = 24.1639; @free(hv4);
hv5 = 8.5931; @free(hv5);
hv6 = 16.3428; @free(hv6);
hv7 = 11.9165; @free(hv7);
hv8 = 10.9079; @free(hv8);
hv9 = 13.6543; @free(hv9);
hv10 = 7.4283; @free(hv10);
hv201 = -0.2279; @free(hv201);
hv202 = 0.8647; @free(hv202);
hv203 = 1.265; @free(hv203);
hv204_05 = -5.0052; @free(hv204_05);
hv207_14 = 0; @free(hv207_14);

!Group contribution for liquid molar volume;
mv1 = 0.0241; @free(mv1);
mv2 = 0.0165; @free(mv2);
mv3 = 0.0086; @free(mv3);
mv4 = 0.0044; @free(mv4);

```



```

mv5 = 0.0228; @free(mv5);
mv6 = 0.0281; @free(mv6);
mv7 = 0.026; @free(mv7);
mv8 = 0.0209; @free(mv8);
mv9 = 0.0259; @free(mv9);
mv10 = 0.0187; @free(mv10);
mv201 = 0.0009; @free(mv201);
mv202 = -0.0021; @free(mv202);
mv203 = -0.0004; @free(mv203);
mv204_05 = 0.0015; @free(mv204_05);
mv207_14 = -0.0006; @free(mv207_14);

!Group contribution for octanol-water partition coefficient
(Kow);
kow1 = 0.3008; @free(kow1);
kow2 = 0.4352; @free(kow2);
kow3 = 0.3837; @free(kow3);
kow4 = -1.0185; @free(kow4);
kow5 = -0.1449; @free(kow5);
kow6 = -1.465; @free(kow6);
kow7 = -0.9465; @free(kow7);
kow8 = -0.4419; @free(kow8);
kow9 = -0.3519; @free(kow9);
kow10 = -0.6373; @free(kow10);
kow201 = 0.1169; @free(kow201);
kow202 = 0.0193; @free(kow202);
kow203 = -0.0449; @free(kow203);
kow204_05 = -0.1041; @free(kow204_05);
kow207_14 = -0.0975; @free(kow207_14);

!Group contribution for flash point (Fp);
fp1 = 21.7458; @free(fp1);
fp2 = 11.5194; @free(fp2);
fp3 = -5.1205; @free(fp3);
fp4 = 78.5878; @free(fp4);
fp5 = 32.914; @free(fp5);
fp6 = 63.0277; @free(fp6);
fp7 = 38.5602; @free(fp7);
fp8 = 12.2191; @free(fp8);
fp9 = 49.4137; @free(fp9);
fp10 = -2.1262; @free(fp10);
fp201 = 0.1812; @free(fp201);
fp202 = 3.5328; @free(fp202);
fp203 = 1.0254; @free(fp203);
fp204_05 = 8.5848; @free(fp204_05);
fp207_14 = 5.0306; @free(fp207_14);

!Group contribution for melting point (Tm);
tm1 = 0.6699; @free(tm1);
tm2 = 0.2992; @free(tm2);
tm3 = -0.2943; @free(tm3);
tm4 = 3.2702; @free(tm4);
tm5 = 0.7649; @free(tm5);
tm6 = 3.4368; @free(tm6);
tm7 = 2.0673; @free(tm7);
tm8 = 1.6571; @free(tm8);
tm9 = 0.9396; @free(tm9);
tm10 = -0.1982; @free(tm10);
tm201 = 0.0426; @free(tm201);
tm202 = 0.164; @free(tm202);
tm203 = -0.0049; @free(tm203);

```

```

tm204_05 = -0.2099; @free(tm204_05);
tm207_14 = -0.1908; @free(tm207_14);

!Group contribution for viscosity;
vis1 = -1.0278; @free(vis1);
vis2 = 0.2125; @free(vis2);
vis3 = 1.318; @free(vis3);
vis4 = 1.3057; @free(vis4);
vis5 = 0.6134; @free(vis5);
vis6 = 0.2902; @free(vis6);
vis7 = 1.0512; @free(vis7);
vis8 = 1.8378; @free(vis8);
vis9 = 0.8715; @free(vis9);
vis10 = 1.4376; @free(vis10);
vis201 = 0.0142; @free(vis201);
vis202 = 0.4075; @free(vis202);
vis203 = -0.2116; @free(vis203);
vis204_05 = 0; @free(vis204_05);
vis207_14 = 0.6128; @free(vis207_14);

!Group contribution for permissible exposure limit (PEL);
pel1 = 0.7723; @free(pel1);
pel2 = 0.0727; @free(pel2);
pel3 = -0.6557; @free(pel3);
pel4 = 1.3612; @free(pel4);
pel5 = 0.9276; @free(pel5);
pel6 = 1.9265; @free(pel6);
pel7 = 1.2126; @free(pel7);
pel8 = 1.2708; @free(pel8);
pel9 = 1.1981; @free(pel9);
pel10 = 0.2724; @free(pel10);
pel201 = -0.0213; @free(pel201);
pel202 = 0.0043; @free(pel202);
pel203 = 0.0954; @free(pel203);
pel204_05 = 0.2618; @free(pel204_05);
pel207_14 = 0.3565; @free(pel207_14);

!Group contribution for molecular weight (MW);
mw1 = 15.035;
mw2 = 14.027;
mw3 = 13.019;
mw4 = 17.007;
mw5 = 30.026;
mw6 = 30.05;
mw7 = 29.042;
mw8 = 28.034;
mw9 = 29.042;
mw10 = 28.034;

!Group contribution for oral rat lethal dosage (LD50);
ld1 = -0.0742; @free(ld1);
ld2 = 0.0223; @free(ld2);
ld3 = 0.1335; @free(ld3);
ld4 = -0.1955; @free(ld4);
ld5 = 0.0974; @free(ld5);
ld6 = 0.045; @free(ld6);
ld7 = 0.2571; @free(ld7);
ld8 = 0.2506; @free(ld8);
ld9 = 0.3338; @free(ld9);
ld10 = 0.4337; @free(ld10);
ld201 = -0.0308; @free(ld201);

```

```

ld202 = 0.0468; @free(ld202);
ld203 = 0.0219; @free(ld203);
ld204_05 = 0.0066; @free(ld204_05);
ld207_14 = -0.0205; @free(ld207_14);

!Property Constraints;
!Equation for normal boiling point;
tb_tot =
n1*tb1+n2*tb2+n3*tb3+n4*tb4+n5*tb5+n6*tb6+n7*tb7+n8*tb8+n9*tb9+
n10*tb10+n201*tb201+n202*tb202+n203*tb203+(n204+n205)*tb204_05
+(n207+n209+n210+n211+n212+n213+n214+n215+n216)*tb207_14;
@free(tb_tot);

!Equation for enthalpy of vaporization;
hv_tot =
n1*hv1+n2*hv2+n3*hv3+n4*hv4+n5*hv5+n6*hv6+n7*hv7+n8*hv8+n9*hv9+
n10*hv10+n201*hv201+n202*hv202+n203*hv203+(n204+n205)*hv204_05
+(n207+n209+n210+n211+n212+n213+n214+n215+n216)*hv207_14;
@free(hv_tot);

!Equation for liquid molar volume;
mv_tot =
n1*mv1+n2*mv2+n3*mv3+n4*mv4+n5*mv5+n6*mv6+n7*mv7+n8*mv8+n9*mv9+
n10*mv10+n201*mv201+n202*mv202+n203*mv203+(n204+n205)*mv204_05
+(n207+n209+n210+n211+n212+n213+n214+n215+n216)*mv207_14;
@free(mv_tot);

!Equation for octanol-water partition coefficient (Kow);
kow_tot =
n1*kow1+n2*kow2+n3*kow3+n4*kow4+n5*kow5+n6*kow6+n7*kow7+n8*kow8
+n9*kow9+n10*kow10+n201*kow201+n202*kow202+n203*kow203+(n204+n2
05)*kow204_05
+(n207+n209+n210+n211+n212+n213+n214+n215+n216)*kow207_14;
@free(kow_tot);

!Equation for flash point (Fp);
fp_tot =
n1*fp1+n2*fp2+n3*fp3+n4*fp4+n5*fp5+n6*fp6+n7*fp7+n8*fp8+n9*fp9+
n10*fp10+n201*fp201+n202*fp202+n203*fp203+(n204+n205)*fp204_05
+(n207+n209+n210+n211+n212+n213+n214+n215+n216)*fp207_14;
@free(fp_tot);

!Equation for normal melting point (Tm);
tm_tot =
n1*tm1+n2*tm2+n3*tm3+n4*tm4+n5*tm5+n6*tm6+n7*tm7+n8*tm8+n9*tm9+
n10*tm10+n201*tm201+n202*tm202+n203*tm203+(n204+n205)*tm204_05
+(n207+n209+n210+n211+n212+n213+n214+n215+n216)*tm207_14;
@free(tm_tot);

!Equation for viscosity;
vis_tot =
n1*vis1+n2*vis2+n3*vis3+n4*vis4+n5*vis5+n6*vis6+n7*vis7+n8*vis8
+n9*vis9+n10*vis10+n201*vis201+n202*vis202+n203*vis203+(n204+n2
05)*vis204_05
+(n207+n209+n210+n211+n212+n213+n214+n215+n216)*vis207_14;
@free(vis_tot);

!Equation for lower flammability limit (LFL);
o2_coef = C_number+H_number/4-O_number/2;
o2_coef > 0.305531;

```

```

!Equation for molecular weight;
mw =
n1*mw1+n2*mw2+n3*mw3+n4*mw4+n5*mw5+n6*mw6+n7*mw7+n8*mw8+n9*mw9+
n10*mw10;

!Equation for permissible exposure limit (PEL);
pel_tot =
n1*pel1+n2*pel2+n3*pel3+n4*pel4+n5*pel5+n6*pel6+n7*pel7+n8*pel8
+n9*pel9+n10*pel10+n201*pel201+n202*pel202+n203*pel203+(n204+n2
05)*pel204_05
+(n207+n209+n210+n211+n212+n213+n214+n215+n216)*pel207_14;
@free(pel_tot);

!Equation for oral rat lethal dosage (LD50);
ld_tot =
n1*ld1+n2*ld2+n3*ld3+n4*ld4+n5*ld5+n6*ld6+n7*ld7+n8*ld8+n9*ld9+
n10*ld10+n201*ld201+n202*ld202+n203*ld203+(n204+n205)*ld204_05
+(n207+n209+n210+n211+n212+n213+n214+n215+n216)*ld207_14;
@free(ld_tot);
ld_tot2 = 0.0016*mw;
ld_tot3 = ld_tot+1.9372+ld_tot2; @free(ld_tot3);
ld_total = ld_tot3-3-@log10(mw); @free(ld_total);

!Properties upper and lower boundaries;
@BND(5.22893, tb_tot, 12.787892);
@BND(4.262302, tm_tot, 7.977927);
@BND(40.3873, hv_tot, 518.3873);
@BND(0.024, mv_tot, 0.208);
@BND(60, mw, 250);
@BND(-2.302585, vis_tot, 4.60517);

!Molecular structure constraints;
n6+n7+n8+n9+n10=1; !the upper limit number of each group;

n5<=1;
(0-0.5)*I_n5 < n5-0.5;
n5-0.5 <= (1-0.5)*(1-I_n5);

n4<=I_n5*3+(1-I_n5); @BIN(I_n5);

!Structural constraints;
!Defining maximum number of structural groups in a molecule;
DATA:
nmax = 13;
ENDDATA

!Defining valence number;
SETS:
Molecular_group_i /1..10/: Valence, gin_i1;
id_no /1..4/;
y_i_id(Molecular_group_i, id_no, Molecular_group_i, id_no):
Binary_y_i_id;
z_i_id(Molecular_group_i, id_no): Binary_z_i_id;
w_i_id(Molecular_group_i, id_no): Binary_w_i_id;
ENDSETS

@FOR(Molecular_group_i: @GIN(gin_i1));
@FOR(y_i_id: @BIN(Binary_y_i_id));
@FOR(z_i_id: @BIN(Binary_z_i_id));
@FOR(w_i_id: @BIN(Binary_w_i_id));

```

```

DATA:
Valence = 1 2 3 1 2 1 2 3 2 3;
ENDDATA

@SUM(Molecular_group_i: (2 - Valence) * gin_il) = 2;
@FOR(Molecular_group_i(i2): @SUM(Molecular_group_i(i1)| i1 #NE#
i2: gin_il(i1)) >= gin_il(i2) * (Valence(i2) - 2) + 2);

@FOR(Molecular_group_i(i1): @FOR(id_no(id1):
Binary_y_i_id(i1,id1,i1,id1) = 0));

@FOR(Molecular_group_i(i1)| i1 #GE# 2: @FOR(id_no(id1)| id1
#GE# 2: @SUM(y_i_id(i1,id1,i2,id2)| i2 #LE# (i1-1):
Binary_y_i_id(i1,id1,i2,id2)) +
@SUM(y_i_id(i1,id1,i1,id2)| id2 #LE# (id1-1):
Binary_y_i_id(i1,id1,i1,id2)) >= -Binary_w_i_id(i1,id1));
@SUM(Molecular_group_i: gin_il) + @SUM(w_i_id: Binary_w_i_id) =
nmax;

@FOR(Molecular_group_i(i)| i #EQ# 1: @FOR(id_no(id)| id #EQ# 1:
Binary_w_i_id(i,id) = 0));
@FOR(Molecular_group_i(i1): @FOR(Molecular_group_i(i2)| i1 #GE#
(i2+1): @FOR(id_no(id1): @FOR(id_no(id2):

Binary_w_i_id(i1,id1) >= Binary_w_i_id(i2,id2)))));
@FOR(Molecular_group_i(i1): @FOR(id_no(id1): @FOR(id_no(id2)|
id1 #GE# (id2+1): Binary_w_i_id(i1,id1) >=
Binary_w_i_id(i1,id2)))));

@FOR(Molecular_group_i(i1): @FOR(Molecular_group_i(i2):
@FOR(id_no(id1): @FOR(id_no(id2):
Binary_y_i_id(i1,id1,i2,id2) =
Binary_y_i_id(i2,id2,i1,id1)))));

@FOR(Molecular_group_i(i1): @FOR(id_no(id1):
@SUM(y_i_id(i1,id1,i2,id2): Binary_y_i_id(i1,id1,i2,id2)) =
Valence(i1)*Binary_z_i_id(i1,id1));

@FOR(Molecular_group_i(i1): @SUM(z_i_id(i1,id1):
Binary_z_i_id(i1,id1) = gin_il(i1));

!Special constraints;
@FOR(Molecular_group_i(i)| i #EQ# 4: @FOR(id_no(id)| id #EQ# 4:
Binary_z_i_id(i,id) = 0));
@FOR(Molecular_group_i(i)| i #GE# 5: @FOR(id_no(id)| id #GE# 2:
Binary_z_i_id(i,id) = 0));

@FOR(Molecular_group_i(i1)| i1 #EQ# 4:
@FOR(Molecular_group_i(i2): @FOR(id_no(id1)| id1 #EQ# 4:
@FOR(id_no(id2): Binary_y_i_id(i1,id1,i2,id2) = 0)))));
@FOR(Molecular_group_i(i1)| i1 #GE# 5:
@FOR(Molecular_group_i(i2): @FOR(id_no(id1)| id1 #GE# 2:
@FOR(id_no(id2): Binary_y_i_id(i1,id1,i2,id2) = 0)))));

!Constraint for CH3;
@FOR(id_no(id1): @FOR(id_no(id2): Binary_y_i_id(1,id1,1,id2) =
0));
@FOR(id_no(id1): @FOR(id_no(id2): Binary_y_i_id(1,id1,4,id2) =
0));

!Constraint for OH and CH3N;

```

```

Binary_y_i_id(4,1,9,1) = 0; Binary_y_i_id(4,2,9,1) = 0;
Binary_y_i_id(4,3,9,1) = 0;
Binary_y_i_id(4,1,4,2) = 0; Binary_y_i_id(4,1,4,3) = 0;
Binary_y_i_id(4,2,4,3) = 0;

!Constraint for CH2NH;
Binary_y_i_id(7,1,4,1) + Binary_y_i_id(7,1,4,2) +
Binary_y_i_id(7,1,4,3) <= 1;
Binary_y_i_id(7,1,1,1) + Binary_y_i_id(7,1,1,2) +
Binary_y_i_id(7,1,1,3) + Binary_y_i_id(7,1,1,4) <= 1;

!Constraint for CHNH;
Binary_y_i_id(8,1,4,1) + Binary_y_i_id(8,1,4,2) +
Binary_y_i_id(8,1,4,3) <= 1;

-0.5*(1-I_n8) < n8-0.5;
n8-0.5 <= 0.5*I_n8; @BIN(I_n8);

Binary_y_i_id(8,1,3,1) + Binary_y_i_id(8,1,3,2) +
Binary_y_i_id(8,1,3,3) + Binary_y_i_id(8,1,3,4) >= I_n8;

!Constraint for CH2N;
Binary_y_i_id(10,1,4,1) + Binary_y_i_id(10,1,4,2) +
Binary_y_i_id(10,1,4,3) <= 1;
Binary_y_i_id(10,1,1,1) + Binary_y_i_id(10,1,1,2) +
Binary_y_i_id(10,1,1,3) + Binary_y_i_id(10,1,1,4) <= 1;

Binary_y_i_id(3,1,3,2)+Binary_y_i_id(3,1,1,1)+Binary_y_i_id(3,1
,1,2)+Binary_y_i_id(3,1,1,3)+Binary_y_i_id(3,1,1,4)+
Binary_y_i_id(3,2,1,1)+Binary_y_i_id(3,2,1,2)+Binary_y_i_id(3,2
,1,3)+Binary_y_i_id(3,2,1,4) <= 4;

Binary_y_i_id(3,1,3,3)+Binary_y_i_id(3,1,1,1)+Binary_y_i_id(3,1
,1,2)+Binary_y_i_id(3,1,1,3)+Binary_y_i_id(3,1,1,4)+
Binary_y_i_id(3,3,1,1)+Binary_y_i_id(3,3,1,2)+Binary_y_i_id(3,3
,1,3)+Binary_y_i_id(3,3,1,4) <= 4;

Binary_y_i_id(3,1,3,4)+Binary_y_i_id(3,1,1,1)+Binary_y_i_id(3,1
,1,2)+Binary_y_i_id(3,1,1,3)+Binary_y_i_id(3,1,1,4)+
Binary_y_i_id(3,4,1,1)+Binary_y_i_id(3,4,1,2)+Binary_y_i_id(3,4
,1,3)+Binary_y_i_id(3,4,1,4) <= 4;

Binary_y_i_id(3,2,3,3)+Binary_y_i_id(3,2,1,1)+Binary_y_i_id(3,2
,1,2)+Binary_y_i_id(3,2,1,3)+Binary_y_i_id(3,2,1,4)+
Binary_y_i_id(3,3,1,1)+Binary_y_i_id(3,3,1,2)+Binary_y_i_id(3,3
,1,3)+Binary_y_i_id(3,3,1,4) <= 4;

Binary_y_i_id(3,2,3,4)+Binary_y_i_id(3,2,1,1)+Binary_y_i_id(3,2
,1,2)+Binary_y_i_id(3,2,1,3)+Binary_y_i_id(3,2,1,4)+
Binary_y_i_id(3,4,1,1)+Binary_y_i_id(3,4,1,2)+Binary_y_i_id(3,4
,1,3)+Binary_y_i_id(3,4,1,4) <= 4;

Binary_y_i_id(3,3,3,4)+Binary_y_i_id(3,3,1,1)+Binary_y_i_id(3,3
,1,2)+Binary_y_i_id(3,3,1,3)+Binary_y_i_id(3,3,1,4)+
Binary_y_i_id(3,4,1,1)+Binary_y_i_id(3,4,1,2)+Binary_y_i_id(3,4
,1,3)+Binary_y_i_id(3,4,1,4) <= 4;

n1 = gin_il(1); n2 = gin_il(2); n3 = gin_il(3); n4 = gin_il(4);
n5 = gin_il(5);
n6 = gin_il(6); n7 = gin_il(7); n8 = gin_il(8); n9 = gin_il(9);
n10 = gin_il(10);

```

!2nd-order groups [only the constraints for (CH3)2CH are shown here, for the constraints of other 2nd-order groups, please contact the authors];

!(CH3)2CH;

```
Binary_y_i_id(3,1,1,1)+Binary_y_i_id(3,1,1,2)+Binary_y_i_id(3,1,1,3)+Binary_y_i_id(3,1,1,4)-1.5 >= -1.5*(1-I_n20101);
Binary_y_i_id(3,1,1,1)+Binary_y_i_id(3,1,1,2)+Binary_y_i_id(3,1,1,3)+Binary_y_i_id(3,1,1,4)-1.5 < 0.5*I_n20101;
```

```
Binary_y_i_id(3,2,1,1)+Binary_y_i_id(3,2,1,2)+Binary_y_i_id(3,2,1,3)+Binary_y_i_id(3,2,1,4)-1.5 >= -1.5*(1-I_n20102);
Binary_y_i_id(3,2,1,1)+Binary_y_i_id(3,2,1,2)+Binary_y_i_id(3,2,1,3)+Binary_y_i_id(3,2,1,4)-1.5 < 0.5*I_n20102;
```

```
Binary_y_i_id(3,3,1,1)+Binary_y_i_id(3,3,1,2)+Binary_y_i_id(3,3,1,3)+Binary_y_i_id(3,3,1,4)-1.5 >= -1.5*(1-I_n20103);
Binary_y_i_id(3,3,1,1)+Binary_y_i_id(3,3,1,2)+Binary_y_i_id(3,3,1,3)+Binary_y_i_id(3,3,1,4)-1.5 < 0.5*I_n20103;
```

```
Binary_y_i_id(3,4,1,1)+Binary_y_i_id(3,4,1,2)+Binary_y_i_id(3,4,1,3)+Binary_y_i_id(3,4,1,4)-1.5 >= -1.5*(1-I_n20104);
Binary_y_i_id(3,4,1,1)+Binary_y_i_id(3,4,1,2)+Binary_y_i_id(3,4,1,3)+Binary_y_i_id(3,4,1,4)-1.5 < 0.5*I_n20104;
```

```
Binary_y_i_id(8,1,1,1)+Binary_y_i_id(8,1,1,2)+Binary_y_i_id(8,1,1,3)+Binary_y_i_id(8,1,1,4)-1.5 >= -1.5*(1-I_n20105);
Binary_y_i_id(8,1,1,1)+Binary_y_i_id(8,1,1,2)+Binary_y_i_id(8,1,1,3)+Binary_y_i_id(8,1,1,4)-1.5 < 1.5*I_n20105;
```

```
Binary_y_i_id(8,1,1,1)+Binary_y_i_id(8,1,1,2)+Binary_y_i_id(8,1,1,3)+Binary_y_i_id(8,1,1,4)-2.5 >= -2.5*(1-I_n20106);
Binary_y_i_id(8,1,1,1)+Binary_y_i_id(8,1,1,2)+Binary_y_i_id(8,1,1,3)+Binary_y_i_id(8,1,1,4)-2.5 < 0.5*I_n20106;
```

```
I_n20107 <= I_n20105;
I_n20107 >= I_n20106;
```

```
n201 = I_n20101+I_n20102+I_n20103+I_n20104+I_n20107;
@BIN(I_n20101); @BIN(I_n20102); @BIN(I_n20103); @BIN(I_n20104);
@BIN(I_n20105); @BIN(I_n20106); @BIN(I_n20107);
```

!Inherent safety and health penalty score;

!Flammability (I\_fl);

!Fp >= 93.4'C or fp\_tot >= 195.8442, penalty score = 1

Fp < 93.4'C or fp\_tot < 195.8442, penalty score = 2

Fp < 37.8'C or fp\_tot < 140.2442, penalty score = 3

Fp < 22.8'C or fp\_tot < 125.2442 and Tb <= 37.8'C or tb\_tot <= 3.566887, penalty score = 4;

!Disjunctive programming algorithm for I\_fl;

(29.4442-195.8442)\*I\_fl1 <= fp\_tot-195.8442;

fp\_tot-195.8442 < (429.4442-195.8442)\*(1-I\_fl1);

(29.4442-140.2442)\*I\_fl2 < fp\_tot-140.2442;

fp\_tot-140.2442 <= (429.4442-140.2442)\*(1-I\_fl2);

(29.4442-125.2442)\*(1-I\_fl3) < fp\_tot-125.2442;

fp\_tot-125.2442 <= (429.4442-125.2442)\*I\_fl3;

```

(1-3.566887)*(1-I_fl4) <= tb_tot-3.566887;
tb_tot-3.566887 < (26.357763-3.566887)*I_fl4;

(0-0.5)*(1-I_fl5) <= I_fl3+I_fl4-0.5;
I_fl3+I_fl4-0.5 < (2-0.5)*I_fl5;

I_fl = 2+I_fl1+I_fl2-I_fl5;
@BIN(I_fl1); @BIN(I_fl2); @BIN(I_fl3); @BIN(I_fl4);
@BIN(I_fl5);

!Material state (I_ms);
!Gas (Tm < 25'C or tm_tot < 7.977927 and Tb < 25'C or tb_tot <
3.38497), penalty score = 1
Liquid (Tm < 25'C or tm_tot < 7.977927 and Tb >= 25'C or tb_tot
>= 3.38497), penalty score = 2
Solid (Tm >= 25'C or tm_tot >= 7.977927 and Tb >= 25'C or
tb_tot >= 3.38497), penalty score = 3;

!Disjunctive programming algorithm for I_ms;
(1-3.38497)*I_ms1 < tb_tot-3.38497;
tb_tot-3.38497 <= (26.357763-3.38497)*(1-I_ms1);

I_ms = 2-I_ms1;
@BIN(I_ms1);

!Volatility (I_v);
!Tb > 150'C or tb_tot > 5.643803, penalty score = 0
150'C >= Tb > 50'C or 5.643803 >= tb_tot > 3.74937, penalty
score = 1
50'C >= Tb > 0'C or 3.74937 >= tb_tot > 3.055987, penalty score
= 2
Tb <= 0'C or tb_tot <= 3.055987, penalty score = 3;

!Disjunctive programming algorithm for I_v;
(1-5.643803)*(1-I_v1) <= tb_tot-5.643803;
tb_tot-5.643803 < (26.357763-5.643803)*I_v1;

(1-3.74937)*(1-I_v2) <= tb_tot-3.74937;
tb_tot-3.74937 < (26.357763-3.74937)*I_v2;

(1-3.055987)*(1-I_v3) <= tb_tot-3.055987;
tb_tot-3.055987 < (26.357763-3.055987)*I_v3;

I_v = 3-I_v1-I_v2-I_v3;
@BIN(I_v1); @BIN(I_v2); @BIN(I_v3);

!Viscosity (I_vis);
!0.1 - 1 cp or vis_tot < 0, penalty score = 1
1 - 10 cp or 0 <= vis_tot < 2.3026, penalty score = 2
10 - 100 cp or vis_tot >= 2.3026, penalty score = 3;

!Disjunctive programming algorithm for I_vis;
(-2.302585-0)*I_vis1 < vis_tot-0;
vis_tot-0 <= (4.60517-0)*(1-I_vis1);

(-2.302585-2.302585)*I_vis2 < vis_tot-2.302585;
vis_tot-2.302585 <= (4.60517-2.302585)*(1-I_vis2);

I_vis = 3-I_vis1-I_vis2;
@BIN(I_vis1); @BIN(I_vis2);

```



```

!Explosive limit (I_ex);
!0 <= ex_range < 20 or o2_coef > 2.679362, penalty score = 1
20 <= ex_range < 45 or 0.482205 < o2_coef <= 2.679362, penalty
score = 2
45 <= ex_range < 70 or o2_coef <= 0.482205, penalty score = 3;

!Disjunctive programming algorithm for I_ex;
(0-2.679362)*(1-I_ex1) < o2_coef-2.679362;
o2_coef-2.679362 <= (100-2.679362)*I_ex1;

(0-0.482205)*(1-I_ex2) < o2_coef-0.482205;
o2_coef-0.482205 <= (100-0.482205)*I_ex2;

I_ex = 3-I_ex1-I_ex2;
@BIN(I_ex1); @BIN(I_ex2);

!Permissible exposure limit (I_el);
!Liquid and Vapour
PEL > 1000 or pel_tot < 1.388279, penalty score = 0
PEL <= 1000 or pel_tot >= 1.388279, penalty score = 1
PEL <= 100 or pel_tot >= 2.388279, penalty score = 2
PEL <= 10 pel_tot >= 3.388279, penalty score = 3
PEL <= 1 pel_tot >= 4.388279, penalty score = 4;

!Disjunctive programming algorithm for I_el;
(7.388279-1.388279)*(1-I_el1) >= pel_tot-1.388279;
pel_tot-1.388279 > (-1.611721-1.388279)*I_el1;

(7.388279-2.388279)*(1-I_el2) >= pel_tot-2.388279;
pel_tot-2.388279 > (-1.611721-2.388279)*I_el2;

(7.388279-3.388279)*(1-I_el3) >= pel_tot-3.388279;
pel_tot-3.388279 > (-1.611721-3.388279)*I_el3;

(7.388279-4.388279)*(1-I_el4) >= pel_tot-4.388279;
pel_tot-4.388279 > (-1.611721-4.388279)*I_el4;

I_el = 4-I_el1-I_el2-I_el3-I_el4;
@BIN(I_el1); @BIN(I_el2); @BIN(I_el3); @BIN(I_el4);

!Acute health hazard (I_ah);
!LD50 > 2000 or ld_total < -3.30103, penalty score = 0
PEL <= 2000 or ld_total >= -3.30103, penalty score = 1
PEL <= 500 or ld_total >= -2.69897, penalty score = 2
PEL <= 50 or ld_total >= -1.69897, penalty score = 3
PEL <= 5 or ld_total >= -0.69897, penalty score = 4;

!Disjunctive programming algorithm for I_ah;
(2.30103+3.30103)*(1-I_ah1) >= ld_total+3.30103;
ld_total+3.30103 > (-6.69897+3.30103)*I_ah1;

(2.30103+2.69897)*(1-I_ah2) >= ld_total+2.69897;
ld_total+2.69897 > (-6.69897+2.69897)*I_ah2;

(2.30103+1.69897)*(1-I_ah3) >= ld_total+1.69897;
ld_total+1.69897 > (-6.69897+1.69897)*I_ah3;

(2.30103+0.69897)*(1-I_ah4) >= ld_total+0.69897;
ld_total+0.69897 > (-6.69897+0.69897)*I_ah4;

I_ah = 4-I_ah1-I_ah2-I_ah3-I_ah4;

```

```

@BIN(I_ah1); @BIN(I_ah2); @BIN(I_ah3); @BIN(I_ah4);

!Summation of sub-index values to calculate total index score;
I_SHI = I_fl+I_ms+I_v+I_vis+I_ex+I_el+I_ah; @GIN(I_SHI);

END

```

## A.3 Lingo Coding (Chapter 6)

```

model:

!Objective function;

max = lambda;

!Linear membership functions for objectives;
(11.80962-I_SHI)/(11.80962-8.767422) = lambda_I;
tb_diff = @sqrt((tb_tot-4.0441)^2);
(2.7261-tb_diff)/(2.7261-0.0032) = lambda_p1;
(30.8018-hv_tot)/(30.8018-14.0244) = lambda_p2;
(9.777867-Ra)/(9.777867-3.586346) = lambda_p3;

lambda_I >= lambda;
lambda_p1 >= lambda;
lambda_p2 >= lambda;
lambda_p3 >= lambda;

lambda >= 0;
lambda <= 1;

!Defining the chemical building blocks
1st-order groups
n1 = CH3, n2 = CH2, n3 = CH, n4 = C, n5 = OH, n6 = CH3CO, n7 =
CH2CO, n8 = CH3O, n9 = CH2O, n10 = CHO, n15 = CH3COO, n16 =
CH2COO, n25 = CH2(cyc), n26 = CH(cyc), n27 = C(cyc), n28 =
O(cyc);

@GIN(n1); @GIN(n2); @GIN(n3); @GIN(n4); @GIN(n5); @GIN(n6);
@GIN(n7); @GIN(n8); @GIN(n9); @GIN(n10); @GIN(n15); @GIN(n16);
@GIN(n25); @GIN(n26); @GIN(n27); @GIN(n28);

n_tot = n1+n2+n3+n4+n5+n6+n7+n8+n9+n10+n15+n16+n25+n26+n27+n28;

C_number =
n1+n2+n3+n4+n6*2+n7*2+n8+n9+n10+n15*2+n16*2+n25+n26+n27;
H_number =
n1*3+n2*2+n3+n5+n6*3+n7*2+n8*3+n9*2+n10+n15*3+n16*2+n25*2+n26;
O_number = n5+n6+n7+n8+n9+n10+n15*2+n16*2+n28;

!Group contribution for normal boiling point (Tb);
tb1 = 0.8853; @free(tb1);
tb2 = 0.5815; @free(tb2);
tb3 = -0.0039; @free(tb3);
tb4 = -0.4985; @free(tb4);
tb5 = 2.1385; @free(tb5);
tb6 = 2.6245; @free(tb6);
tb7 = 2.0151; @free(tb7);
tb8 = 1.5724; @free(tb8);
tb9 = 0.9999; @free(tb9);

```

```

tb10 = 0.4724; @free(tb10);
tb15 = 2.5805; @free(tb15);
tb16 = 2.1808; @free(tb16);
tb25 = 0.7067; @free(tb25);
tb26 = 0.3922; @free(tb26);
tb27 = -0.2034; @free(tb27);
tb28 = 0.8691; @free(tb28);

!Group contribution for normal melting point (Tm);
tm1 = 0.6699; @free(tm1);
tm2 = 0.2992; @free(tm2);
tm3 = -0.2943; @free(tm3);
tm4 = -0.043; @free(tm4);
tm5 = 3.2702; @free(tm5);
tm6 = 3.1357; @free(tm6);
tm7 = 2.9007; @free(tm7);
tm8 = 1.5327; @free(tm8);
tm9 = 0.7649; @free(tm9);
tm10 = 0.1817; @free(tm10);
tm15 = 2.4227; @free(tm15);
tm16 = 1.5439; @free(tm16);
tm25 = 0.5067; @free(tm25);
tm26 = 0.2691; @free(tm26);
tm27 = 0.5775; @free(tm27);
tm28 = 1.3269; @free(tm28);

!Group contribution for octanol-water partition coefficient
(log Kow);
kow1 = 0.3008; @free(kow1);
kow2 = 0.4352; @free(kow2);
kow3 = 0.3837; @free(kow3);
kow4 = 0.6325; @free(kow4);
kow5 = -1.0185; @free(kow5);
kow6 = -0.3774; @free(kow6);
kow7 = -0.1855; @free(kow7);
kow8 = -0.303; @free(kow8);
kow9 = -0.1449; @free(kow9);
kow10 = 0.165; @free(kow10);
kow15 = -0.4615; @free(kow15);
kow16 = -0.2893; @free(kow16);
kow25 = 0.1818; @free(kow25);
kow26 = 0.2934; @free(kow26);
kow27 = 0.2412; @free(kow27);
kow28 = -0.4008; @free(kow28);

!Group contribution for flash point (Fp);
fp1 = 21.7458; @free(fp1);
fp2 = 11.5194; @free(fp2);
fp3 = -5.1205; @free(fp3);
fp4 = -19.7535; @free(fp4);
fp5 = 78.5878; @free(fp5);
fp6 = 70.9382; @free(fp6);
fp7 = 67.479; @free(fp7);
fp8 = 41.9635; @free(fp8);
fp9 = 32.914; @free(fp9);
fp10 = -8.9309; @free(fp10);
fp15 = 73.7009; @free(fp15);
fp16 = 50.2088; @free(fp16);
fp25 = 15.0958; @free(fp25);
fp26 = 10.5355; @free(fp26);
fp27 = -15.1444; @free(fp27);

```

```

fp28 = 24.3976; @free(fp28);

!Group contribution for enthalpy of vaporization at Tb (Hv);
hv1 = 2.0701;
hv2 = 2.3353;
hv3 = 1.6963;
hv4 = 0.8251;
hv5 = 16.4887;
hv6 = 9.619;
hv7 = 10.1585;
hv8 = 5.6432;
hv9 = 5.1006;
hv10 = 4.3695;
hv15 = 13.0656;
hv16 = 13.5635;
hv25 = 2.2775;
hv26 = 2.4151;
hv27 = 2.9031;
hv28 = 4.7567;

!Group contribution for Hansen solubility parameter -
dispersion (deld);
deld1 = 7.5697; @free(deld1);
deld2 = -0.0018; @free(deld2);
deld3 = -7.7208; @free(deld3);
deld4 = -15.4498; @free(deld4);
deld5 = 8.0236; @free(deld5);
deld6 = 8.163; @free(deld6);
deld7 = 0.5557; @free(deld7);
deld8 = 7.6577; @free(deld8);
deld9 = 0.1978; @free(deld9);
deld10 = -7.7099; @free(deld10);
deld15 = 8.022; @free(deld15);
deld16 = 0.4586; @free(deld16);
deld25 = 2.6915; @free(deld25);
deld26 = -3.7719; @free(deld26);
deld27 = -7.187; @free(deld27);
deld28 = 3.9616; @free(deld28);

!Group contribution for Hansen solubility parameter - polar
(delp);
delp1 = 1.9996; @free(delp1);
delp2 = -0.1492; @free(delp2);
delp3 = -2.7099; @free(delp3);
delp4 = -4.7191; @free(delp4);
delp5 = 4.9598; @free(delp5);
delp6 = 6.052; @free(delp6);
delp7 = 0.7632; @free(delp7);
delp8 = 3.086; @free(delp8);
delp9 = 0.6423; @free(delp9);
delp10 = -1.918; @free(delp10);
delp15 = 2.848; @free(delp15);
delp16 = 1.4477; @free(delp16);
delp25 = 0.5026; @free(delp25);
delp26 = -1.7549; @free(delp26);
delp27 = -2.2674; @free(delp27);
delp28 = 3.1902; @free(delp28);

!Group contribution for Hansen solubility parameter - H2 bond
(delh);
delh1 = 2.2105; @free(delh1);

```

```

delh2 = -0.215; @free(delh2);
delh3 = -2.6826; @free(delh3);
delh4 = -6.4821; @free(delh4);
delh5 = 11.8005; @free(delh5);
delh6 = 3.4394; @free(delh6);
delh7 = -0.0788; @free(delh7);
delh8 = 3.3464; @free(delh8);
delh9 = 0.8246; @free(delh9);
delh10 = -2.1543; @free(delh10);
delh15 = 5.0132; @free(delh15);
delh16 = 2.7824; @free(delh16);
delh25 = 0.6159; @free(delh25);
delh26 = -0.5171; @free(delh26);
delh27 = -2.6329; @free(delh27);
delh28 = 2.802; @free(delh28);

!Group contribution for liquid molar volume (Vm);
mv1 = 0.0241;
mv2 = 0.0165;
mv3 = 0.0086;
mv4 = 0.0007;
mv5 = 0.0044;
mv6 = 0.0345;
mv7 = 0.0288;
mv8 = 0.0283;
mv9 = 0.0228;
mv10 = 0.0207;
mv15 = 0.0412;
mv16 = 0.0365;
mv25 = 0.0159;
mv26 = 0.0063;
mv27 = 0.0006;
mv28 = 0.0018;

!Group contribution for lower flammability limit (LFL);
lfl1 = -0.2357; @free(lfl1);
lfl2 = -0.2334; @free(lfl2);
lfl3 = -0.2308; @free(lfl3);
lfl4 = -0.2161; @free(lfl4);
lfl5 = 0.0599; @free(lfl5);
lfl6 = -0.3205; @free(lfl6);
lfl7 = -0.1764; @free(lfl7);
lfl8 = -0.1921; @free(lfl8);
lfl9 = -0.1213; @free(lfl9);
lfl10 = -0.2958; @free(lfl10);
lfl15 = -0.2264; @free(lfl15);
lfl16 = -0.6266; @free(lfl16);
lfl25 = -0.2169; @free(lfl25);
lfl26 = -0.2941; @free(lfl26);
lfl27 = -0.1401; @free(lfl27);
lfl28 = 0.1086; @free(lfl28);

!Group contribution for upper flammability limit (UFL);
ufl1 = -1.1534; @free(ufl1);
ufl2 = -0.1445; @free(ufl2);
ufl3 = 0.8856; @free(ufl3);
ufl4 = 1.8649; @free(ufl4);
ufl5 = -0.7578; @free(ufl5);
ufl6 = -1.1643; @free(ufl6);
ufl7 = -0.171; @free(ufl7);
ufl8 = -0.8561; @free(ufl8);

```

```

ufl9 = 0.2096; @free(ufl9);
ufl10 = 0.9939; @free(ufl10);
ufl15 = -1.2311; @free(ufl15);
ufl16 = 0.077; @free(ufl16);
ufl25 = -0.4403; @free(ufl25);
ufl26 = 2.0503; @free(ufl26);
ufl27 = 1.0217; @free(ufl27);
ufl28 = -0.0295; @free(ufl28);

!Group contribution for viscosity;
vis1 = -1.0278; @free(vis1);
vis2 = 0.2125; @free(vis2);
vis3 = 1.318; @free(vis3);
vis4 = 2.8147; @free(vis4);
vis5 = 1.3057; @free(vis5);
vis6 = -0.1881; @free(vis6);
vis7 = 0.9647; @free(vis7);
vis8 = -0.6902; @free(vis8);
vis9 = 0.6134; @free(vis9);
vis10 = 3.6344; @free(vis10);
vis15 = -0.0358; @free(vis15);
vis16 = 1.0292; @free(vis16);
vis25 = -0.0577; @free(vis25);
vis26 = 0.9455; @free(vis26);
vis27 = 1.5824; @free(vis27);
vis28 = 0.0434; @free(vis28);

!Group contribution for fathead minnow 96-h LC50;
lc1 = 0.6172; @free(lc1);
lc2 = 0.4464; @free(lc2);
lc3 = 0.1522; @free(lc3);
lc4 = -0.1861; @free(lc4);
lc5 = -0.2125; @free(lc5);
lc6 = 0.6176; @free(lc6);
lc7 = 0.4468; @free(lc7);
lc8 = 0.378; @free(lc8);
lc9 = 0.2072; @free(lc9);
lc10 = -0.087; @free(lc10);
lc15 = 1.5633; @free(lc15);
lc16 = 1.3925; @free(lc16);
lc25 = 0.4464; @free(lc25);
lc26 = 0.1522; @free(lc26);
lc27 = -0.1861; @free(lc27);
lc28 = -0.2392; @free(lc28);

!Group contribution for oral rat lethal dosage (LD50);
ld1 = -0.0742; @free(ld1);
ld2 = 0.0223; @free(ld2);
ld3 = 0.1335; @free(ld3);
ld4 = 0.2641; @free(ld4);
ld5 = -0.1955; @free(ld5);
ld6 = -0.0172; @free(ld6);
ld7 = 0.1931; @free(ld7);
ld8 = -0.0259; @free(ld8);
ld9 = 0.0974; @free(ld9);
ld10 = 0.4987; @free(ld10);
ld15 = -0.1734; @free(ld15);
ld16 = -0.0357; @free(ld16);
ld25 = 0.0305; @free(ld25);
ld26 = 0.1009; @free(ld26);
ld27 = 0.2675; @free(ld27);

```

```

ld28 = 0.0485; @free(ld28);

!Group contribution for permissible exposure limit (PEL);
pel1 = 0.7723; @free(pel1);
pel2 = 0.0727; @free(pel2);
pel3 = -0.6557; @free(pel3);
pel4 = -1.3404; @free(pel4);
pel5 = 1.3612; @free(pel5);
pel6 = 1.4016; @free(pel6);
pel7 = 1.2601; @free(pel7);
pel8 = 2.1251; @free(pel8);
pel9 = 0.9276; @free(pel9);
pel10 = -0.7462; @free(pel10);
pel15 = 1.2544; @free(pel15);
pel16 = 1.6798; @free(pel16);
pel25 = 0.2678; @free(pel25);
pel26 = -0.1033; @free(pel26);
pel27 = -0.6719; @free(pel27);
pel28 = 1.0976; @free(pel28);

!Group contribution for molecular weight (MW);
mw1 = 15.035;
mw2 = 14.027;
mw3 = 13.019;
mw4 = 12.011;
mw5 = 17.007;
mw6 = 43.045;
mw7 = 42.037;
mw8 = 31.034;
mw9 = 30.026;
mw10 = 29.018;
mw15 = 59.044;
mw16 = 58.036;
mw25 = 14.027;
mw26 = 13.019;
mw27 = 12.011;
mw28 = 15.999;

!Property Constraints;
!Equation for normal boiling point (Tb);
tb_tot =
n1*tb1+n2*tb2+n3*tb3+n4*tb4+n5*tb5+n6*tb6+n7*tb7+n8*tb8+n9*tb9+
n10*tb10+n15*tb15+n16*tb16+n25*tb25+n26*tb26+n27*tb27+n28*tb28;
@free(tb_tot);
tb = 244.5165*@log(tb_tot)-273.15; @free(tb);

!Equation for normal melting point (Tm);
tm_tot =
n1*tm1+n2*tm2+n3*tm3+n4*tm4+n5*tm5+n6*tm6+n7*tm7+n8*tm8+n9*tm9+
n10*tm10+n15*tm15+n16*tm16+n25*tm25+n26*tm26+n27*tm27+n28*tm28;
@free(tm_tot);
tm = 143.5706*@log(tm_tot)-273.15; @free(tm);

!Equation for octanol-water partition coefficient (log Kow);
kow_tot =
n1*kow1+n2*kow2+n3*kow3+n4*kow4+n5*kow5+n6*kow6+n7*kow7+n8*kow8
+n9*kow9+n10*kow10+n15*kow15+n16*kow16+n25*kow25+n26*kow26+n27*
kow27+n28*kow28; @free(kow_tot);
log_kow = kow_tot+0.4876; @free(log_kow);

!Equation for flash point (Fp);

```

```

fp_tot =
n1*fp1+n2*fp2+n3*fp3+n4*fp4+n5*fp5+n6*fp6+n7*fp7+n8*fp8+n9*fp9+
n10*fp10+n15*fp15+n16*fp16+n25*fp25+n26*fp26+n27*fp27+n28*fp28;
@free(fp_tot);
fp = fp_tot+170.7058-273.15; @free(fp);

!Equation for enthalpy of vaporization at Tb (Hv);
hv_tot =
n1*hv1+n2*hv2+n3*hv3+n4*hv4+n5*hv5+n6*hv6+n7*hv7+n8*hv8+n9*hv9+
n10*hv10+n15*hv15+n16*hv16+n25*hv25+n26*hv26+n27*hv27+n28*hv28;
hvb = hv_tot+15.4199;

!Equation for Hansen solubility parameter - dispersion (deld);
deld =
n1*deld1+n2*deld2+n3*deld3+n4*deld4+n5*deld5+n6*deld6+n7*deld7+
n8*deld8+n9*deld9+n10*deld10+n15*deld15+n16*deld16+n25*deld25+n
26*deld26+n27*deld27+n28*deld28; @free(deld);

!Equation for Hansen solubility parameter - polar (delp);
delp =
n1*delp1+n2*delp2+n3*delp3+n4*delp4+n5*delp5+n6*delp6+n7*delp7+
n8*delp8+n9*delp9+n10*delp10+n15*delp15+n16*delp16+n25*delp25+n
26*delp26+n27*delp27+n28*delp28; @free(delp);

!Equation for Hansen solubility parameter - H2 bond (delh);
delh =
n1*delh1+n2*delh2+n3*delh3+n4*delh4+n5*delh5+n6*delh6+n7*delh7+
n8*delh8+n9*delh9+n10*delh10+n15*delh15+n16*delh16+n25*delh25+n
26*delh26+n27*delh27+n28*delh28; @free(delh);

!Equation for liquid molar volume;
mv_tot =
n1*mv1+n2*mv2+n3*mv3+n4*mv4+n5*mv5+n6*mv6+n7*mv7+n8*mv8+n9*mv9+
n10*mv10+n15*mv15+n16*mv16+n25*mv25+n26*mv26+n27*mv27+n28*mv28;
mv = mv_tot+0.016;

!Equation for lower flammability limit (LFL);
lfl_tot =
n1*lfl1+n2*lfl2+n3*lfl3+n4*lfl4+n5*lfl5+n6*lfl6+n7*lfl7+n8*lfl8
+n9*lfl9+n10*lfl10+n15*lfl15+n16*lfl16+n25*lfl25+n26*lfl26+n27*
lfl27+n28*lfl28; @free(lfl_tot);
lfl = 4.5315*@exp(lfl_tot);

!Equation for upper flammability limit (UFL);
ufl_tot =
n1*ufl1+n2*ufl2+n3*ufl3+n4*ufl4+n5*ufl5+n6*ufl6+n7*ufl7+n8*ufl8
+n9*ufl9+n10*ufl10+n15*ufl15+n16*ufl16+n25*ufl25+n26*ufl26+n27*
ufl27+n28*ufl28; @free(ufl_tot);
ufl = 129.9552*@exp(ufl_tot);
ex_range = ufl-lfl;

!Equation for viscosity;
vis_tot =
n1*vis1+n2*vis2+n3*vis3+n4*vis4+n5*vis5+n6*vis6+n7*vis7+n8*vis8
+n9*vis9+n10*vis10+n15*vis15+n16*vis16+n25*vis25+n26*vis26+n27*
vis27+n28*vis28; @free(vis_tot);
vis = @exp(vis_tot);

!Equation for molecular weight;

```



```

mw =
n1*mw1+n2*mw2+n3*mw3+n4*mw4+n5*mw5+n6*mw6+n7*mw7+n8*mw8+n9*mw9+
n10*mw10+n15*mw15+n16*mw16+n25*mw25+n26*mw26+n27*mw27+n28*mw28;

!Equation for permissible exposure limit (PEL);
pel_tot =
n1*pel1+n2*pel2+n3*pel3+n4*pel4+n5*pel5+n6*pel6+n7*pel7+n8*pel8
+n9*pel9+n10*pel10+n15*pel15+n16*pel16+n25*pel25+n26*pel26+n27*
pel27+n28*pel28; @free(pel_tot);

!Equation for oral rat lethal dosage (LD50);
ld_tot =
n1*ld1+n2*ld2+n3*ld3+n4*ld4+n5*ld5+n6*ld6+n7*ld7+n8*ld8+n9*ld9+
n10*ld10+n15*ld15+n16*ld16+n25*ld25+n26*ld26+n27*ld27+n28*ld28;
@free(ld_tot);
ld_tot2 = 0.0016*mw;
ld_tot3 = ld_tot+1.9372+ld_tot2; @free(ld_tot3);
ld_total = ld_tot3-3-@log10(mw); @free(ld_total);

!Equation for fathead minnow 96-h LC50;
lc_tot =
n1*lc1+n2*lc2+n3*lc3+n4*lc4+n5*lc5+n6*lc6+n7*lc7+n8*lc8+n9*lc9+
n10*lc10+n15*lc15+n16*lc16+n25*lc25+n26*lc26+n27*lc27+n28*lc28;
@free(lc_tot);
lc_tot2 = lc_tot-I_diol*0.4639-2*I_diester*(n2+n16+n25)*0.1393;
@free(lc_tot2);
lc_total = lc_tot2-3-@log10(mw); @free(lc_total);
(0-1.5)*(1-I_diol) < n5-1.5;
n5-1.5 <= (2-1.5)*I_diol; @BIN(I_diol);
(0-1.5)*(1-I_diester) < n15+n16-1.5;
n15+n16-1.5 <= (2-1.5)*I_diester; @BIN(I_diester);

!Distance of solvent from solubility sphere (Ra) calculation;
Ra = @sqrt(4*(deld-17.3782)^2+(delp-0.3839)^2+(delh-1.6396)^2);

!Properties upper and lower boundaries;
log_kow < 2.86;
fp >= -1.2;
lc_total < -1.37;
tb_tot <= 4.7914;
Ra <= 7.76107;

!Molecular structure constraints;
n1+n2+n3+n4+n5+n6+n7+n8+n9+n10+n15+n16+n25+n26+n27+n28>0; !a
molecule must be formed;
n5+n6+n7+n8+n9+n10+n15+n16<=2; !groups containing O-atom cannot
appear more than two times;

!Structural constraints to differentiate acyclic and cyclic
compounds;
n25+n26+n27+n28 <= 30*I_cyc; @BIN(I_cyc);
n25+n26+n27+n28 >= 3*I_cyc;
FBN = 2*I_cyc;
n1+n2+n3+n4+n5+n6+n7+n8+n9+n10+n15+n16 <= 30*I_cycl1;
n1+n2+n3+n4+n5+n6+n7+n8+n9+n10+n15+n16 >= I_cycl1; @BIN(I_cycl1);
n26+n27 <= 30*(I_cyc+I_cycl1-1);
n26+n27 >= I_cyc+I_cycl1-1;

!Free bonds for each group;
val1 = 1;
val2 = 2;

```

```

val3 = 3;
val4 = 4;
val5 = 1;
val6 = 1;
val7 = 2;
val8 = 1;
val9 = 2;
val10 = 3;
val11 = 1;
val12 = 2;
val13 = 2;
val14 = 3;
val15 = 1;
val16 = 2;
val17 = 1;
val18 = 1;
val19 = 1;
val20 = 2;
val21 = 1;
val22 = 2;
val23 = 2;
val24 = 3;
val25 = 2;
val26 = 3;
val27 = 4;
val28 = 2;

!Structural constraint, the molecule generated must not contain
free bonds;
(n1*val1+n2*val2+n3*val3+n4*val4+n5*val5+n6*val6+n7*val7+n8*val
8+n9*val9+n10*val10+n11*val11+n12*val12+n13*val13+n14*val14+n15
*val15+n16*val16+n17*val17+
n18*val18+n19*val19+n20*val20+n21*val21+n22*val22+n23*val23+n24
*val24+n25*val25+n26*val26+n27*val27+n28*val28)-
(2*(n1+n2+n3+n4+n5+n6+n7+n8+n9+n10+n11+n12+
n13+n14+n15+n16+n17+n18+n19+n20+n21+n22+n23+n24+n25+n26+n27+n28
-1)) = FBN;

!Inherent safety and health penalty score;
!Flammability (I_fl);
!Fp >= 93.4'C or fp_tot >= 195.8442, penalty score = 1
Fp < 93.4'C or fp_tot < 195.8442, penalty score = 2
Fp < 37.8'C or fp_tot < 140.2442, penalty score = 3
Fp < 22.8'C or fp_tot < 125.2442 and Tb <= 37.8'C or tb_tot <=
3.566887, penalty score = 4;

!Disjunctive programming algorithm for I_fl;
(29.4442-195.8442)*I_fl1 <= fp_tot-195.8442;
fp_tot-195.8442 < (429.4442-195.8442)*(1-I_fl1);

(29.4442-140.2442)*I_fl2 < fp_tot-140.2442;
fp_tot-140.2442 <= (429.4442-140.2442)*(1-I_fl2);

(29.4442-125.2442)*(1-I_fl3) < fp_tot-125.2442;
fp_tot-125.2442 <= (429.4442-125.2442)*I_fl3;

(1-3.566887)*(1-I_fl4) <= tb_tot-3.566887;
tb_tot-3.566887 < (26.357763-3.566887)*I_fl4;

(0-0.5)*(1-I_fl5) <= I_fl3+I_fl4-0.5;
I_fl3+I_fl4-0.5 < (2-0.5)*I_fl5;

```

```

I_fl = 2+I_fl1+I_fl2-I_fl5;
@BIN(I_fl1); @BIN(I_fl2); @BIN(I_fl3); @BIN(I_fl4);
@BIN(I_fl5);

!Material state (I_ms);
!Gas (Tb < 17.1'C or tb_tot < 3.277354), penalty score = 1
(17.1'C <= Tb < 32.9'C or 3.277354 <= tb_tot < 3.49612),
penalty score = (tb_tot-3.277354)/0.218766+1
Liquid (Tb >= 32.9'C or tb_tot >= 3.49612), penalty score = 2;

!Disjunctive programming algorithm for I_ms;
(-273.15-17.1)*(1-I_ms1) < tb-17.1;
tb-17.1 <= (526.85-17.1)*I_ms1;

(-273.15-32.9)*(1-I_ms2) < tb-32.9;
tb-32.9 <= (526.85-32.9)*I_ms2;

I_ms = 1+(I_ms1-I_ms2)*(tb-17.1)/15.8+I_ms2;
@BIN(I_ms1); @BIN(I_ms2);

!Volatility (I_v);
!Tb > 157.9'C or tb_tot > 5.829124, penalty score = 0
157.9'C >= Tb > 142.1'C or 5.829124 >= tb_tot > 5.464373,
penalty score = (5.829124-tb_tot)/0.364751
142.1'C >= Tb > 57.9'C or 5.464373 >= tb_tot > 3.872485,
penalty score = 1
57.9'C >= Tb > 42.1'C or 3.872485 >= tb_tot > 3.630168, penalty
score = (3.872485-tb_tot)/0.242317+1
42.1'C >= Tb > 7.9'C or 3.630168 >= tb_tot > 3.156334, penalty
score = 2
7.9'C >= Tb > -7.9'C or 3.156334 >= tb_tot > 2.95883, penalty
score = (3.156334-tb_tot)/0.197504+2
Tb <= -7.9'C or tb_tot <= 2.95883, penalty score = 3;

!Disjunctive programming algorithm for I_v;
(-273.15-157.9)*I_v1 <= tb-157.9;
tb-157.9 < (526.85-157.9)*(1-I_v1);

(-273.15-142.1)*I_v2 <= tb-142.1;
tb-142.1 < (526.85-142.1)*(1-I_v2);

(-273.15-57.9)*I_v3 <= tb-57.9;
tb-57.9 < (526.85-57.9)*(1-I_v3);

(-273.15-42.1)*I_v4 <= tb-42.1;
tb-42.1 < (526.85-42.1)*(1-I_v4);

(-273.15-7.9)*I_v5 <= tb-7.9;
tb-7.9 < (526.85-7.9)*(1-I_v5);

(-273.15+7.9)*I_v6 <= tb+7.9;
tb+7.9 < (526.85+7.9)*(1-I_v6);

I_v = (I_v1-I_v2)*(157.9-tb)/15.8+I_v2+(I_v3-I_v4)*(57.9-
tb)/15.8+I_v4+(I_v5-I_v6)*(7.9-tb)/15.8+I_v6;
@BIN(I_v1); @BIN(I_v2); @BIN(I_v3); @BIN(I_v4); @BIN(I_v5);
@BIN(I_v6);

!Viscosity (I_vis);
!0.1 - 1 cp, penalty score = (vis-0.1)/1.8+1

```

```

1 - 4.44 cp, penalty score = (vis-1)/6.88+1.5
4.44 - 6.56 cp, penalty score = 2
6.56 - 13.44 cp, penalty score = (vis-6.56)/6.88+2
13.44 - 100 cp, penalty score = 3;

!Disjunctive programming algorithm for I_vis;
(0.1-1)*(1-I_vis1) < vis-1;
vis-1 <= (100-1)*I_vis1;

(0.1-4.44)*(1-I_vis2) < vis-4.44;
vis-4.44 <= (100-4.44)*I_vis2;

(0.1-6.56)*(1-I_vis3) < vis-6.56;
vis-6.56 <= (100-6.56)*I_vis3;

(0.1-13.44)*(1-I_vis4) < vis-13.44;
vis-13.44 <= (100-13.44)*I_vis4;

I_vis = 1+(1-I_vis1)*(vis-0.1)/1.8+(I_vis1-I_vis2)*((vis-1)/6.88+0.5)+I_vis2+(I_vis3-I_vis4)*(vis-6.56)/6.88+I_vis4;
@BIN(I_vis1); @BIN(I_vis2); @BIN(I_vis3); @BIN(I_vis4);

!Explosive limit (I_ex);
!0 <= ex_range < 20, penalty score = 1
20 <= ex_range < 45, penalty score = 2
45 <= ex_range < 70, penalty score = 3
70 <= ex_range <= 100, penalty score = 4;

!Disjunctive programming algorithm for I_ex;
(0-20)*(1-I_ex1) < ex_range-20;
ex_range-20 <= (100-20)*I_ex1;

(0-45)*(1-I_ex2) < ex_range-45;
ex_range-45 <= (100-45)*I_ex2;

(0-70)*(1-I_ex3) < ex_range-70;
ex_range-70 <= (100-70)*I_ex3;

I_ex = 1+I_ex1+I_ex2+I_ex3;
@BIN(I_ex1); @BIN(I_ex2); @BIN(I_ex3);

!Permissible exposure limit (I_el);
!Liquid and Vapour
pel_tot < 0.608279, penalty score = 0
0.608279 <= pel_tot < 1.608279, penalty score = (pel_tot-0.608279)/1.56
1.608279 <= pel_tot < 2.168279, penalty score = (pel_tot-1.608279)/0.56*28/39+25/39
2.168279 <= pel_tot < 2.608279, penalty score = (pel_tot-1.608279)/1.56+1
2.608279 <= pel_tot < 3.168279, penalty score = (pel_tot-2.608279)/0.56*28/39+64/39
3.168279 <= pel_tot < 3.608279, penalty score = (pel_tot-2.608279)/1.56+2
3.608279 <= pel_tot < 4.168279, penalty score = (pel_tot-3.608279)/0.56*28/39+103/39
4.168279 <= pel_tot < 5.168279, penalty score = (pel_tot-3.608279)/1.56+3
pel_tot >= 5.168279, penalty score = 4;

!Disjunctive programming algorithm for I_el;

```

```

(-1.611721-0.608279)*(1-I_el1) < pel_tot-0.608279;
pel_tot-0.608279 <= (7.388279-0.608279)*I_el1;

(-1.611721-1.608279)*(1-I_el2) < pel_tot-1.608279;
pel_tot-1.608279 <= (7.388279-1.608279)*I_el2;

(-1.611721-2.168279)*(1-I_el3) < pel_tot-2.168279;
pel_tot-2.168279 <= (7.388279-2.168279)*I_el3;

(-1.611721-2.608279)*(1-I_el4) < pel_tot-2.608279;
pel_tot-2.608279 <= (7.388279-2.608279)*I_el4;

(-1.611721-3.168279)*(1-I_el5) < pel_tot-3.168279;
pel_tot-3.168279 <= (7.388279-3.168279)*I_el5;

(-1.611721-3.608279)*(1-I_el6) < pel_tot-3.608279;
pel_tot-3.608279 <= (7.388279-3.608279)*I_el6;

(-1.611721-4.168279)*(1-I_el7) < pel_tot-4.168279;
pel_tot-4.168279 <= (7.388279-4.168279)*I_el7;

(-1.611721-5.168279)*(1-I_el8) < pel_tot-5.168279;
pel_tot-5.168279 <= (7.388279-5.168279)*I_el8;

I_el = (I_el1-I_el2)*(pel_tot-0.608279)/1.56+(I_el2-
I_el3)*(pel_tot-1.608279)/0.56*28/39+25/39)+(I_el3-
I_el4)*(pel_tot-1.608279)/1.56+1)
+(I_el4-I_el5)*(pel_tot-2.608279)/0.56*28/39+64/39)+(I_el5-
I_el6)*(pel_tot-2.608279)/1.56+2)+(I_el6-I_el7)*(pel_tot-
3.608279)/0.56*28/39+103/39)
+(I_el7-I_el8)*(pel_tot-3.608279)/1.56+3)+I_el8*4;
@BIN(I_el1); @BIN(I_el2); @BIN(I_el3); @BIN(I_el4);
@BIN(I_el5); @BIN(I_el6); @BIN(I_el7); @BIN(I_el8);

!Acute health hazard (I_ah);
!ld_total < -3.73103, penalty score = 0
-3.73103 <= ld_total < -3.12897, penalty score =
(ld_total+3.73103)/0.86
-3.12897 <= ld_total < -2.87103, penalty score =
(ld_total+3.12897)/0.25794*12897/21500+30103/43000
-2.87103 <= ld_total < -2.26897, penalty score =
(ld_total+3.12897)/0.86+1
-2.26897 <= ld_total < -2.12897, penalty score = 2
-2.12897 <= ld_total < -1.26897, penalty score =
(ld_total+2.12897)/0.86+2
-1.26897 <= ld_total < -1.12897, penalty score = 3
-1.12897 <= ld_total < -0.26897, penalty score =
(ld_total+1.12897)/0.86+3
ld_total >= -0.26897, penalty score = 4;

!Disjunctive programming algorithm for I_ah;
(-6.69897+3.73103)*(1-I_ah1) < ld_total+3.73103;
ld_total+3.73103 <= (2.30103+3.73103)*I_ah1;

(-6.69897+3.12897)*(1-I_ah2) < ld_total+3.12897;
ld_total+3.12897 <= (2.30103+3.12897)*I_ah2;

(-6.69897+2.87103)*(1-I_ah3) < ld_total+2.87103;
ld_total+2.87103 <= (2.30103+2.87103)*I_ah3;

(-6.69897+2.26897)*(1-I_ah4) < ld_total+2.26897;

```

```

ld_total+2.26897 <= (2.30103+2.26897)*I_ah4;

(-6.69897+2.12897)*(1-I_ah5) < ld_total+2.12897;
ld_total+2.12897 <= (2.30103+2.12897)*I_ah5;

(-6.69897+1.26897)*(1-I_ah6) < ld_total+1.26897;
ld_total+1.26897 <= (2.30103+1.26897)*I_ah6;

(-6.69897+1.12897)*(1-I_ah7) < ld_total+1.12897;
ld_total+1.12897 <= (2.30103+1.12897)*I_ah7;

(-6.69897+0.26897)*(1-I_ah8) < ld_total+0.26897;
ld_total+0.26897 <= (2.30103+0.26897)*I_ah8;

I_ah = (I_ah1-I_ah2)*(ld_total+3.73103)/0.86+(I_ah2-
I_ah3)*((ld_total+3.12897)/0.25794*12897/21500+30103/43000)+(I_
ah3-I_ah4)*((ld_total+3.12897)/0.86+1)
+(I_ah4-I_ah5)*2+(I_ah5-
I_ah6)*((ld_total+2.12897)/0.86+2)+(I_ah6-I_ah7)*3+(I_ah7-
I_ah8)*((ld_total+1.12897)/0.86+3)+I_ah8*4;
@BIN(I_ah1); @BIN(I_ah2); @BIN(I_ah3); @BIN(I_ah4);
@BIN(I_ah5); @BIN(I_ah6); @BIN(I_ah7); @BIN(I_ah8);

!Summation of sub-index values to calculate total index score;
I_SHI = I_fl+I_ms+I_v+I_vis+I_ex+I_el+I_ah;

END

```

## A.4 Lingo Coding (Chapter 7)

```

model:

!Objective function;

max = lambda;

!Linear membership functions for objectives;
(2.655308-I_SHIr)/(2.655308-1.746318) = lambda_I;
(7.3889-tb_tot)/(7.3889-3.0238) = lambda_p1;
(36.88-hv_tot)/(36.88-11.6891) = lambda_p2;
(15.89908-Ra)/(15.89908-4.05417) = lambda_p3;

lambda_I >= lambda;
lambda_p1 >= lambda;
lambda_p2 >= lambda;
lambda_p3 >= lambda;

lambda >= 0;
lambda <= 1;

!Defining the chemical building blocks
1st-order groups
n1 = CH3, n2 = CH2, n3 = CH, n4 = C, n5 = OH, n6 = CH3CO, n7 =
CH2CO, n8 = CH3O, n9 = CH2O, n10 = CHO
n11 = CH3COO, n12 = CH2COO, n13 = CH2(cyc), n14 = CH(cyc), n15
= C(cyc), n16 = O(cyc);

@GIN(n1); @GIN(n2); @GIN(n3); @GIN(n4); @GIN(n5); @GIN(n6);
@GIN(n7); @GIN(n8); @GIN(n9); @GIN(n10);

```

```

@GIN(n11); @GIN(n12); @GIN(n13); @GIN(n14); @GIN(n15);
@GIN(n16);

n_tot = n1+n2+n3+n4+n5+n6+n7+n8+n9+n10+n11+n12+n13+n14+n15+n16;

C_number =
n1+n2+n3+n4+n6*2+n7*2+n8+n9+n10+n11*2+n12*2+n13+n14+n15;
H_number =
n1*3+n2*2+n3+n5+n6*3+n7*2+n8*3+n9*2+n10+n11*3+n12*2+n13*2+n14;
O_number = n5+n6+n7+n8+n9+n10+n11*2+n12*2+n16;

!Group contribution for normal boiling point (Tb);
tb1 = 0.8853; @free(tb1);
tb2 = 0.5815; @free(tb2);
tb3 = -0.0039; @free(tb3);
tb4 = -0.4985; @free(tb4);
tb5 = 2.1385; @free(tb5);
tb6 = 2.6245; @free(tb6);
tb7 = 2.0151; @free(tb7);
tb8 = 1.5724; @free(tb8);
tb9 = 0.9999; @free(tb9);
tb10 = 0.4724; @free(tb10);
tb11 = 2.5805; @free(tb11);
tb12 = 2.1808; @free(tb12);
tb13 = 0.7067; @free(tb13);
tb14 = 0.3922; @free(tb14);
tb15 = -0.2034; @free(tb15);
tb16 = 0.8691; @free(tb16);

!Group contribution for normal melting point (Tm);
tm1 = 0.6699; @free(tm1);
tm2 = 0.2992; @free(tm2);
tm3 = -0.2943; @free(tm3);
tm4 = -0.043; @free(tm4);
tm5 = 3.2702; @free(tm5);
tm6 = 3.1357; @free(tm6);
tm7 = 2.9007; @free(tm7);
tm8 = 1.5327; @free(tm8);
tm9 = 0.7649; @free(tm9);
tm10 = 0.1817; @free(tm10);
tm11 = 2.4227; @free(tm11);
tm12 = 1.5439; @free(tm12);
tm13 = 0.5067; @free(tm13);
tm14 = 0.2691; @free(tm14);
tm15 = 0.5775; @free(tm15);
tm16 = 1.3269; @free(tm16);

!Group contribution for octanol-water partition coefficient
(log Kow);
kow1 = 0.3008; @free(kow1);
kow2 = 0.4352; @free(kow2);
kow3 = 0.3837; @free(kow3);
kow4 = 0.6325; @free(kow4);
kow5 = -1.0185; @free(kow5);
kow6 = -0.3774; @free(kow6);
kow7 = -0.1855; @free(kow7);
kow8 = -0.303; @free(kow8);
kow9 = -0.1449; @free(kow9);
kow10 = 0.165; @free(kow10);
kow11 = -0.4615; @free(kow11);
kow12 = -0.2893; @free(kow12);

```

```

kow13 = 0.1818; @free(kow13);
kow14 = 0.2934; @free(kow14);
kow15 = 0.2412; @free(kow15);
kow16 = -0.4008; @free(kow16);

!Group contribution for flash point (Fp);
fp1 = 21.7458; @free(fp1);
fp2 = 11.5194; @free(fp2);
fp3 = -5.1205; @free(fp3);
fp4 = -19.7535; @free(fp4);
fp5 = 78.5878; @free(fp5);
fp6 = 70.9382; @free(fp6);
fp7 = 67.479; @free(fp7);
fp8 = 41.9635; @free(fp8);
fp9 = 32.914; @free(fp9);
fp10 = -8.9309; @free(fp10);
fp11 = 73.7009; @free(fp11);
fp12 = 50.2088; @free(fp12);
fp13 = 15.0958; @free(fp13);
fp14 = 10.5355; @free(fp14);
fp15 = -15.1444; @free(fp15);
fp16 = 24.3976; @free(fp16);

!Group contribution for enthalpy of vaporization at Tb (Hv);
hv1 = 2.0701;
hv2 = 2.3353;
hv3 = 1.6963;
hv4 = 0.8251;
hv5 = 16.4887;
hv6 = 9.619;
hv7 = 10.1585;
hv8 = 5.6432;
hv9 = 5.1006;
hv10 = 4.3695;
hv11 = 13.0656;
hv12 = 13.5635;
hv13 = 2.2775;
hv14 = 2.4151;
hv15 = 2.9031;
hv16 = 4.7567;

!Group contribution for Hansen solubility parameter -
dispersion (deld);
deld1 = 7.5697; @free(deld1);
deld2 = -0.0018; @free(deld2);
deld3 = -7.7208; @free(deld3);
deld4 = -15.4498; @free(deld4);
deld5 = 8.0236; @free(deld5);
deld6 = 8.163; @free(deld6);
deld7 = 0.5557; @free(deld7);
deld8 = 7.6577; @free(deld8);
deld9 = 0.1978; @free(deld9);
deld10 = -7.7099; @free(deld10);
deld11 = 8.022; @free(deld11);
deld12 = 0.4586; @free(deld12);
deld13 = 2.6915; @free(deld13);
deld14 = -3.7719; @free(deld14);
deld15 = -7.187; @free(deld15);
deld16 = 3.9616; @free(deld16);

```



```

!Group contribution for Hansen solubility parameter - polar
(delp);
delp1 = 1.9996; @free(delp1);
delp2 = -0.1492; @free(delp2);
delp3 = -2.7099; @free(delp3);
delp4 = -4.7191; @free(delp4);
delp5 = 4.9598; @free(delp5);
delp6 = 6.052; @free(delp6);
delp7 = 0.7632; @free(delp7);
delp8 = 3.086; @free(delp8);
delp9 = 0.6423; @free(delp9);
delp10 = -1.918; @free(delp10);
delp11 = 2.848; @free(delp11);
delp12 = 1.4477; @free(delp12);
delp13 = 0.5026; @free(delp13);
delp14 = -1.7549; @free(delp14);
delp15 = -2.2674; @free(delp15);
delp16 = 3.1902; @free(delp16);

!Group contribution for Hansen solubility parameter - H2 bond
(delh);
delh1 = 2.2105; @free(delh1);
delh2 = -0.215; @free(delh2);
delh3 = -2.6826; @free(delh3);
delh4 = -6.4821; @free(delh4);
delh5 = 11.8005; @free(delh5);
delh6 = 3.4394; @free(delh6);
delh7 = -0.0788; @free(delh7);
delh8 = 3.3464; @free(delh8);
delh9 = 0.8246; @free(delh9);
delh10 = -2.1543; @free(delh10);
delh11 = 5.0132; @free(delh11);
delh12 = 2.7824; @free(delh12);
delh13 = 0.6159; @free(delh13);
delh14 = -0.5171; @free(delh14);
delh15 = -2.6329; @free(delh15);
delh16 = 2.802; @free(delh16);

!Group contribution for liquid molar volume (Vm);
mv1 = 0.0241;
mv2 = 0.0165;
mv3 = 0.0086;
mv4 = 0.0007;
mv5 = 0.0044;
mv6 = 0.0345;
mv7 = 0.0288;
mv8 = 0.0283;
mv9 = 0.0228;
mv10 = 0.0207;
mv11 = 0.0412;
mv12 = 0.0365;
mv13 = 0.0159;
mv14 = 0.0063;
mv15 = 0.0006;
mv16 = 0.0018;

!Group contribution for lower flammability limit (LFL);
lfl1 = -0.2357; @free(lfl1);
lfl2 = -0.2334; @free(lfl2);
lfl3 = -0.2308; @free(lfl3);
lfl4 = -0.2161; @free(lfl4);

```

```

lfl15 = 0.0599; @free(lfl15);
lfl16 = -0.3205; @free(lfl16);
lfl17 = -0.1764; @free(lfl17);
lfl18 = -0.1921; @free(lfl18);
lfl19 = -0.1213; @free(lfl19);
lfl110 = -0.2958; @free(lfl110);
lfl111 = -0.2264; @free(lfl111);
lfl112 = -0.6266; @free(lfl112);
lfl113 = -0.2169; @free(lfl113);
lfl114 = -0.2941; @free(lfl114);
lfl115 = -0.1401; @free(lfl115);
lfl116 = 0.1086; @free(lfl116);

!Group contribution for upper flammability limit (UFL);
ufl1 = -1.1534; @free(ufl1);
ufl2 = -0.1445; @free(ufl2);
ufl3 = 0.8856; @free(ufl3);
ufl4 = 1.8649; @free(ufl4);
ufl5 = -0.7578; @free(ufl5);
ufl6 = -1.1643; @free(ufl6);
ufl7 = -0.171; @free(ufl7);
ufl8 = -0.8561; @free(ufl8);
ufl9 = 0.2096; @free(ufl9);
ufl110 = 0.9939; @free(ufl110);
ufl111 = -1.2311; @free(ufl111);
ufl112 = 0.077; @free(ufl112);
ufl113 = -0.4403; @free(ufl113);
ufl114 = 2.0503; @free(ufl114);
ufl115 = 1.0217; @free(ufl115);
ufl116 = -0.0295; @free(ufl116);

!Group contribution for viscosity;
vis1 = -1.0278; @free(vis1);
vis2 = 0.2125; @free(vis2);
vis3 = 1.318; @free(vis3);
vis4 = 2.8147; @free(vis4);
vis5 = 1.3057; @free(vis5);
vis6 = -0.1881; @free(vis6);
vis7 = 0.9647; @free(vis7);
vis8 = -0.6902; @free(vis8);
vis9 = 0.6134; @free(vis9);
vis10 = 3.6344; @free(vis10);
vis11 = -0.0358; @free(vis11);
vis12 = 1.0292; @free(vis12);
vis13 = -0.0577; @free(vis13);
vis14 = 0.9455; @free(vis14);
vis15 = 1.5824; @free(vis15);
vis16 = 0.0434; @free(vis16);

!Group contribution for fathead minnow 96-h LC50;
lc1 = 0.6172; @free(lc1);
lc2 = 0.4464; @free(lc2);
lc3 = 0.1522; @free(lc3);
lc4 = -0.1861; @free(lc4);
lc5 = -0.2125; @free(lc5);
lc6 = 0.6176; @free(lc6);
lc7 = 0.4468; @free(lc7);
lc8 = 0.378; @free(lc8);
lc9 = 0.2072; @free(lc9);
lc10 = -0.087; @free(lc10);
lc11 = 1.5633; @free(lc11);

```

```

lc12 = 1.3925; @free(lc12);
lc13 = 0.4464; @free(lc13);
lc14 = 0.1522; @free(lc14);
lc15 = -0.1861; @free(lc15);
lc16 = -0.2392; @free(lc16);

!Group contribution for oral rat lethal dosage (LD50);
ld1 = -0.0742; @free(ld1);
ld2 = 0.0223; @free(ld2);
ld3 = 0.1335; @free(ld3);
ld4 = 0.2641; @free(ld4);
ld5 = -0.1955; @free(ld5);
ld6 = -0.0172; @free(ld6);
ld7 = 0.1931; @free(ld7);
ld8 = -0.0259; @free(ld8);
ld9 = 0.0974; @free(ld9);
ld10 = 0.4987; @free(ld10);
ld11 = -0.1734; @free(ld11);
ld12 = -0.0357; @free(ld12);
ld13 = 0.0305; @free(ld13);
ld14 = 0.1009; @free(ld14);
ld15 = 0.2675; @free(ld15);
ld16 = 0.0485; @free(ld16);

!Group contribution for permissible exposure limit (PEL);
pel1 = 0.7723; @free(pel1);
pel2 = 0.0727; @free(pel2);
pel3 = -0.6557; @free(pel3);
pel4 = -1.3404; @free(pel4);
pel5 = 1.3612; @free(pel5);
pel6 = 1.4016; @free(pel6);
pel7 = 1.2601; @free(pel7);
pel8 = 2.1251; @free(pel8);
pel9 = 0.9276; @free(pel9);
pel10 = -0.7462; @free(pel10);
pel11 = 1.2544; @free(pel11);
pel12 = 1.6798; @free(pel12);
pel13 = 0.2678; @free(pel13);
pel14 = -0.1033; @free(pel14);
pel15 = -0.6719; @free(pel15);
pel16 = 1.0976; @free(pel16);

!Group contribution for molecular weight (MW);
mw1 = 15.035;
mw2 = 14.027;
mw3 = 13.019;
mw4 = 12.011;
mw5 = 17.007;
mw6 = 43.045;
mw7 = 42.037;
mw8 = 31.034;
mw9 = 30.026;
mw10 = 29.018;
mw11 = 59.044;
mw12 = 58.036;
mw13 = 14.027;
mw14 = 13.019;
mw15 = 12.011;
mw16 = 15.999;

!Property Constraints;

```

```

!Equation for normal boiling point (Tb);
tb_tot =
n1*tb1+n2*tb2+n3*tb3+n4*tb4+n5*tb5+n6*tb6+n7*tb7+n8*tb8+n9*tb9+
n10*tb10+n11*tb11+n12*tb12+n13*tb13+n14*tb14+n15*tb15+n16*tb16;
@free(tb_tot);
tb = 244.5165*@log(tb_tot)-273.15; @free(tb);

!Equation for normal melting point (Tm);
tm_tot =
n1*tm1+n2*tm2+n3*tm3+n4*tm4+n5*tm5+n6*tm6+n7*tm7+n8*tm8+n9*tm9+
n10*tm10+n11*tm11+n12*tm12+n13*tm13+n14*tm14+n15*tm15+n16*tm16;
@free(tm_tot);
tm = 143.5706*@log(tm_tot)-273.15; @free(tm);

!Equation for octanol-water partition coefficient (log Kow);
kow_tot =
n1*kow1+n2*kow2+n3*kow3+n4*kow4+n5*kow5+n6*kow6+n7*kow7+n8*kow8+
n9*kow9+n10*kow10+
n11*kow11+n12*kow12+n13*kow13+n14*kow14+n15*kow15+n16*kow16;
@free(kow_tot);
log_kow = kow_tot+0.4876; @free(log_kow);

!Equation for flash point (Fp);
fp_tot =
n1*fp1+n2*fp2+n3*fp3+n4*fp4+n5*fp5+n6*fp6+n7*fp7+n8*fp8+n9*fp9+
n10*fp10+n11*fp11+n12*fp12+n13*fp13+n14*fp14+n15*fp15+n16*fp16;
@free(fp_tot);
fp = fp_tot+170.7058-273.15; @free(fp);

!Equation for enthalpy of vaporization at Tb (Hv);
hv_tot =
n1*hv1+n2*hv2+n3*hv3+n4*hv4+n5*hv5+n6*hv6+n7*hv7+n8*hv8+n9*hv9+
n10*hv10+n11*hv11+n12*hv12+n13*hv13+n14*hv14+n15*hv15+n16*hv16;
hvb = hv_tot+15.4199;

!Equation for Hansen solubility parameter - dispersion (deld);
deld =
n1*deld1+n2*deld2+n3*deld3+n4*deld4+n5*deld5+n6*deld6+n7*deld7+
n8*deld8+n9*deld9+n10*deld10+
n11*deld11+n12*deld12+n13*deld13+n14*deld14+n15*deld15+n16*deld
16; @free(deld);

!Equation for Hansen solubility parameter - polar (delp);
delp =
n1*delp1+n2*delp2+n3*delp3+n4*delp4+n5*delp5+n6*delp6+n7*delp7+
n8*delp8+n9*delp9+n10*delp10+
n11*delp11+n12*delp12+n13*delp13+n14*delp14+n15*delp15+n16*delp
16; @free(delp);

!Equation for Hansen solubility parameter - H2 bond (delh);
delh =
n1*delh1+n2*delh2+n3*delh3+n4*delh4+n5*delh5+n6*delh6+n7*delh7+
n8*delh8+n9*delh9+n10*delh10+
n11*delh11+n12*delh12+n13*delh13+n14*delh14+n15*delh15+n16*delh
16; @free(delh);

!Equation for liquid molar volume;
mv_tot =
n1*mv1+n2*mv2+n3*mv3+n4*mv4+n5*mv5+n6*mv6+n7*mv7+n8*mv8+n9*mv9+
n10*mv10+n11*mv11+n12*mv12+n13*mv13+n14*mv14+n15*mv15+n16*mv16;
mv = mv_tot+0.016;

```

```

!Equation for lower flammability limit (LFL);
lfl_tot =
n1*lfl1+n2*lfl2+n3*lfl3+n4*lfl4+n5*lfl5+n6*lfl6+n7*lfl7+n8*lfl8
+n9*lfl9+n10*lfl10+
n11*lfl11+n12*lfl12+n13*lfl13+n14*lfl14+n15*lfl15+n16*lfl16;
@free(lfl_tot);
lfl = 4.5315*@exp(lfl_tot);

!Equation for upper flammability limit (UFL);
ufl_tot =
n1*ufl1+n2*ufl2+n3*ufl3+n4*ufl4+n5*ufl5+n6*ufl6+n7*ufl7+n8*ufl8
+n9*ufl9+n10*ufl10+
n11*ufl11+n12*ufl12+n13*ufl13+n14*ufl14+n15*ufl15+n16*ufl16;
@free(ufl_tot);
ufl = 129.9552*@exp(ufl_tot);
ex_range = ufl-lfl;

!Equation for viscosity;
vis_tot =
n1*vis1+n2*vis2+n3*vis3+n4*vis4+n5*vis5+n6*vis6+n7*vis7+n8*vis8
+n9*vis9+n10*vis10+
n11*vis11+n12*vis12+n13*vis13+n14*vis14+n15*vis15+n16*vis16;
@free(vis_tot);
vis = vis_tot/2.302585093; @free(vis);
viscosity = @exp(vis_tot);

!Equation for molecular weight;
mw =
n1*mw1+n2*mw2+n3*mw3+n4*mw4+n5*mw5+n6*mw6+n7*mw7+n8*mw8+n9*mw9+
n10*mw10+n11*mw11+n12*mw12+n13*mw13+n14*mw14+n15*mw15+n16*mw16;

!Equation for permissible exposure limit (PEL);
pel_tot =
n1*pel1+n2*pel2+n3*pel3+n4*pel4+n5*pel5+n6*pel6+n7*pel7+n8*pel8
+n9*pel9+n10*pel10+
n11*pel11+n12*pel12+n13*pel13+n14*pel14+n15*pel15+n16*pel16;
@free(pel_tot);

!Equation for oral rat lethal dosage (LD50);
ld_tot =
n1*ld1+n2*ld2+n3*ld3+n4*ld4+n5*ld5+n6*ld6+n7*ld7+n8*ld8+n9*ld9+
n10*ld10+n11*ld11+n12*ld12+n13*ld13+n14*ld14+n15*ld15+n16*ld16;
@free(ld_tot);
ld_tot2 = 0.0016*mw;
ld_tot3 = ld_tot+1.9372+ld_tot2; @free(ld_tot3);
ld_total = ld_tot3-3-@log10(mw); @free(ld_total);

!Equation for fathead minnow 96-h LC50;
lc_tot =
n1*lc1+n2*lc2+n3*lc3+n4*lc4+n5*lc5+n6*lc6+n7*lc7+n8*lc8+n9*lc9+
n10*lc10+n11*lc11+n12*lc12+n13*lc13+n14*lc14+n15*lc15+n16*lc16;
@free(lc_tot);
lc_tot2 = lc_tot-I_diol*0.4639-2*I_diester*(n2+n12+n13)*0.1393;
@free(lc_tot2);
lc_total = lc_tot2-3-@log10(mw); @free(lc_total);
(0-1.5)*(1-I_diol) < n5-1.5;
n5-1.5 <= (2-1.5)*I_diol; @BIN(I_diol);
(0-1.5)*(1-I_diester) < n11+n12-1.5;
n11+n12-1.5 <= (2-1.5)*I_diester; @BIN(I_diester);

```

```

!Distance of solvent from solubility sphere (Ra) calculation;
Ra = @sqrt(4*(deld-17.3782)^2+(delp-0.3839)^2+(delh-1.6396)^2);

!Properties upper and lower boundaries;
fp >= -13.3;
lc_total < -2;

!Molecular structure constraints;
n1+n2+n3+n4+n5+n6+n7+n8+n9+n10+n11+n12+n13+n14+n15+n16>0; !a
molecule must be formed;
n5+n6+n7+n8+n9+n10+n11+n12<=2; !groups containing O-atom cannot
appear more than two times;

!Structural constraints to differentiate acyclic and cyclic
compounds;
n13+n14+n15+n16 <= 30*I_cyc; @BIN(I_cyc);
n13+n14+n15+n16 >= 3*I_cyc;
FBN = 2*I_cyc;
n1+n2+n3+n4+n5+n6+n7+n8+n9+n10+n11+n12 <= 30*I_cycl;
n1+n2+n3+n4+n5+n6+n7+n8+n9+n10+n11+n12 >= I_cycl; @BIN(I_cycl);
n14+n15 <= 30*(I_cyc+I_cycl-1);
n14+n15 >= I_cyc+I_cycl-1;

!Free bonds for each group;
val1 = 1;
val2 = 2;
val3 = 3;
val4 = 4;
val5 = 1;
val6 = 1;
val7 = 2;
val8 = 1;
val9 = 2;
val10 = 3;
val11 = 1;
val12 = 2;
val13 = 2;
val14 = 3;
val15 = 4;
val16 = 2;

!Structural constraint, the molecule generated must not contain
free bonds;
(n1*val1+n2*val2+n3*val3+n4*val4+n5*val5+n6*val6+n7*val7+n8*val
8+n9*val9+n10*val10+n11*val11+n12*val12+n13*val13+n14*val14+n15
*val15+n16*val16)-
(2*(n1+n2+n3+n4+n5+n6+n7+n8+n9+n10+n11+n12+n13+n14+n15+n16-1))
= FBN;

!Inherent safety and health penalty score;
!Flammability (I_fl);
!Fp >= 102.74'C, I_fl_low = I_fl_up = 1
84.06 <= Fp < 102.74'C, I_fl_low = I_fl_up = (102.74-
Fp)/18.68+1
47.14 <= Fp < 84.06'C, I_fl_low = I_fl_up = 2
32.14 <= Fp < 47.14'C, I_fl_low = I_fl_up = (47.14-Fp)/18.68+2
28.46 <= Fp < 32.14'C, I_fl_low = 0.394004283*(32.14-
Fp)/3.68+2.802997859, I_fl_up = (47.14-Fp)/18.68+2
13.46 <= Fp < 28.46'C, I_fl_low = (32.14-Fp)/18.68+3, I_fl_up =
3
Fp < 13.46'C, I_fl_low = 4, I_fl_up = 3

```

```

Tb >= 41.58'C, I_fl = I_fl_up
34.02 <= Tb < 41.58'C, I_fl = (Tb-34.02)/7.56*(I_fl_up-
I_fl_low)+I_fl_low
Tb < 34.02'C, I_fl = I_fl_low;

!Disjunctive programming algorithm for I_fl;
(-73.15-102.74)*I_fl1 < fp-102.74;
fp-102.74 <= (326.85-102.74)*(1-I_fl1);

(-73.15-84.06)*I_fl2 < fp-84.06;
fp-84.06 <= (326.85-84.06)*(1-I_fl2);

(-73.15-47.14)*I_fl3 < fp-47.14;
fp-47.14 <= (326.85-47.14)*(1-I_fl3);

(-73.15-32.14)*I_fl4 < fp-32.14;
fp-32.14 <= (326.85-32.14)*(1-I_fl4);

(-73.15-28.46)*I_fl5 < fp-28.46;
fp-28.46 <= (326.85-28.46)*(1-I_fl5);

(-73.15-13.46)*I_fl6 < fp-13.46;
fp-13.46 <= (326.85-13.46)*(1-I_fl6);

(-273.15-41.58)*I_fl7 < tb-41.58;
tb-41.58 <= (526.85-41.58)*(1-I_fl7);

(-273.15-34.02)*I_fl8 < tb-34.02;
tb-34.02 <= (526.85-34.02)*(1-I_fl8);

I_fl_low = 1+(I_fl1-I_fl2)*(102.74-fp)/18.68+I_fl2-
I_fl3+(I_fl3-I_fl4)*((47.14-fp)/18.68+1)+
(I_fl4-I_fl5)*(0.394004283*(32.14-fp)/3.68+1.802997859)+(I_fl5-
I_fl6)*((32.14-fp)/18.68+2)+I_fl6*3;

I_fl_up = 1+(I_fl1-I_fl2)*(102.74-fp)/18.68+I_fl2-I_fl3+(I_fl3-
I_fl4)*((47.14-fp)/18.68+1)+
(I_fl4-I_fl5)*((47.14-fp)/18.68+1)+I_fl5*2;

I_fl = (1-I_fl7)*I_fl_up+(I_fl7-I_fl8)*((tb-
34.02)/7.56*(I_fl_up-I_fl_low)+I_fl_low)+I_fl8*I_fl_low;

@BIN(I_fl1); @BIN(I_fl2); @BIN(I_fl3); @BIN(I_fl4);
@BIN(I_fl5); @BIN(I_fl6); @BIN(I_fl7); @BIN(I_fl8);

!Material state (I_ms);
!Gas (Tm < 25'C or tm_tot < 7.977927385 and Tb < 25'C or tb_tot
< 3.384970244), penalty score = 1
Liquid (Tm < 25'C or tm_tot < 7.977927385 and Tb >= 25'C or
tb_tot >= 3.384970244), penalty score = 2
Solid (Tm >= 25'C or tm_tot >= 7.977927385 and Tb >= 25'C or
tb_tot >= 3.384970244), penalty score = 3;

!Disjunctive programming algorithm for I_ms;
(1-3.384970244)*(1-I_ms1) < tb_tot-3.384970244;
tb_tot-3.384970244 <= (26.3577631-3.384970244)*I_ms1;

(1-7.977927385)*(1-I_ms2) < tm_tot-7.977927385;
tm_tot-7.977927385 <= (131.0593178-7.977927385)*I_ms2;

I_ms = 1+I_ms1+I_ms2;

```

```

@BIN(I_ms1); @BIN(I_ms2);

!Volatility (I_v);
!Tb > 165'C, penalty score = 0
165'C >= Tb > 135'C, penalty score = (165-tb)/30
135'C >= Tb > 65'C, penalty score = 1
65'C >= Tb > 35'C, penalty score = (65-tb)/30+1
35'C >= Tb > 15'C, penalty score = 2
15'C >= Tb > -15'C, penalty score = (15-tb)/30+2
Tb <= -15'C, penalty score = 3;

!Disjunctive programming algorithm for I_v;
(-273.15-165)*I_v1 <= tb-165;
tb-165 < (526.85-165)*(1-I_v1);

(-273.15-135)*I_v2 <= tb-135;
tb-135 < (526.85-135)*(1-I_v2);

(-273.15-65)*I_v3 <= tb-65;
tb-65 < (526.85-65)*(1-I_v3);

(-273.15-35)*I_v4 <= tb-35;
tb-35 < (526.85-35)*(1-I_v4);

(-273.15-15)*I_v5 <= tb-15;
tb-15 < (526.85-15)*(1-I_v5);

(-273.15+15)*I_v6 <= tb+15;
tb+15 < (526.85+15)*(1-I_v6);

I_v = (I_v1-I_v2)*(165-tb)/30+I_v2+(I_v3-I_v4)*(65-
tb)/30+I_v4+(I_v5-I_v6)*(15-tb)/30+I_v6;
@BIN(I_v1); @BIN(I_v2); @BIN(I_v3); @BIN(I_v4); @BIN(I_v5);
@BIN(I_v6);

!Viscosity (I_vis);
!vis = -1 - -0.1 or 0.1 - 0.7943 cp, penalty score = 1
vis = -0.1 - 0.1 or 0.7943 - 1.2589 cp, penalty score =
(vis+0.1)/0.2+1
vis = 0.1 - 0.9 or 1.2589 - 7.9433 cp, penalty score = 2
vis = 0.9 - 1.1 or 7.9433 - 12.5893 cp, penalty score = (vis-
0.9)/0.2+2
vis = 1.1 - 2 or 12.5893 - 100 cp, penalty score = 3;

!Disjunctive programming algorithm for I_vis;
(-1+0.1)*(1-I_vis1) < vis+0.1;
vis+0.1 <= (2+0.1)*I_vis1;

(-1-0.1)*(1-I_vis2) < vis-0.1;
vis-0.1 <= (2-0.1)*I_vis2;

(-1-0.9)*(1-I_vis3) < vis-0.9;
vis-0.9 <= (2-0.9)*I_vis3;

(-1-1.1)*(1-I_vis4) < vis-1.1;
vis-1.1 <= (2-1.1)*I_vis4;

I_vis = 1+(I_vis1-I_vis2)*(vis+0.1)/0.2+I_vis2+(I_vis3-
I_vis4)*(vis-0.9)/0.2+I_vis4;
@BIN(I_vis1); @BIN(I_vis2); @BIN(I_vis3); @BIN(I_vis4);

```



```

!Explosive limit (I_ex);
!0 <= ex_range < 13, penalty score = 1
13 <= ex_range < 27, penalty score = (ex_range-13)/14+1
27 <= ex_range < 38, penalty score = 2
38 <= ex_range < 52, penalty score = (ex_range-38)/14+2
52 <= ex_range < 63, penalty score = 3
63 <= ex_range < 77, penalty score = (ex_range-63)/14+3
77 <= ex_range <= 100, penalty score = 4;
!Disjunctive programming algorithm for I_ex;
(0-13)*(1-I_ex1) < ex_range-13;
ex_range-13 <= (100-13)*I_ex1;

(0-27)*(1-I_ex2) < ex_range-27;
ex_range-27 <= (100-27)*I_ex2;

(0-38)*(1-I_ex3) < ex_range-38;
ex_range-38 <= (100-38)*I_ex3;

(0-52)*(1-I_ex4) < ex_range-52;
ex_range-52 <= (100-52)*I_ex4;

(0-63)*(1-I_ex5) < ex_range-63;
ex_range-63 <= (100-63)*I_ex5;

(0-77)*(1-I_ex6) < ex_range-77;
ex_range-77 <= (100-77)*I_ex6;

I_ex = 1+(I_ex1-I_ex2)*(ex_range-13)/14+I_ex2+(I_ex3-
I_ex4)*(ex_range-38)/14+I_ex4+(I_ex5-I_ex6)*(ex_range-
63)/14+I_ex6;
@BIN(I_ex1); @BIN(I_ex2); @BIN(I_ex3); @BIN(I_ex4);
@BIN(I_ex5); @BIN(I_ex6);

!Permissible exposure limit (I_el);
!Liquid and Vapour
pel_tot < 1.088278863, penalty score = 0
1.088278863 <= pel_tot < 1.688278863, penalty score = (pel_tot-
1.088278863)/0.6
1.688278863 <= pel_tot < 2.088278863, penalty score = 1
2.088278863 <= pel_tot < 2.688278863, penalty score = (pel_tot-
2.088278863)/0.6+1
2.688278863 <= pel_tot < 3.088278863, penalty score = 2
3.088278863 <= pel_tot < 3.688278863, penalty score = (pel_tot-
3.088278863)/0.6+2
3.688278863 <= pel_tot < 4.088278863, penalty score = 3
4.088278863 <= pel_tot < 4.688278863, penalty score = (pel_tot-
4.088278863)/0.6+3
pel_tot >= 4.688278863, penalty score = 4;

!Disjunctive programming algorithm for I_el;
(-1.611721137-1.088278863)*(1-I_el1) < pel_tot-1.088278863;
pel_tot-1.088278863 <= (7.388278863-1.088278863)*I_el1;

(-1.611721137-1.688278863)*(1-I_el2) < pel_tot-1.688278863;
pel_tot-1.688278863 <= (7.388278863-1.688278863)*I_el2;

(-1.611721137-2.088278863)*(1-I_el3) < pel_tot-2.088278863;
pel_tot-2.088278863 <= (7.388278863-2.088278863)*I_el3;

(-1.611721137-2.688278863)*(1-I_el4) < pel_tot-2.688278863;
pel_tot-2.688278863 <= (7.388278863-2.688278863)*I_el4;

```

```

(-1.611721137-3.088278863)*(1-I_el5) < pel_tot-3.088278863;
pel_tot-3.088278863 <= (7.388278863-3.088278863)*I_el5;

(-1.611721137-3.688278863)*(1-I_el6) < pel_tot-3.688278863;
pel_tot-3.688278863 <= (7.388278863-3.688278863)*I_el6;

(-1.611721137-4.088278863)*(1-I_el7) < pel_tot-4.088278863;
pel_tot-4.088278863 <= (7.388278863-4.088278863)*I_el7;

(-1.611721137-4.688278863)*(1-I_el8) < pel_tot-4.688278863;
pel_tot-4.688278863 <= (7.388278863-4.688278863)*I_el8;

I_el = (I_el1-I_el2)*(pel_tot-1.088278863)/0.6+I_el2+(I_el3-
I_el4)*(pel_tot-2.088278863)/0.6+I_el4
+(I_el5-I_el6)*(pel_tot-3.088278863)/0.6+I_el6+(I_el7-
I_el8)*(pel_tot-4.088278863)/0.6+I_el8;
@BIN(I_el1); @BIN(I_el2); @BIN(I_el3); @BIN(I_el4);
@BIN(I_el5); @BIN(I_el6); @BIN(I_el7); @BIN(I_el8);

!Acute health hazard (I_ah);
!ld_total < -3.631132995, penalty score = 0
-3.631132995 <= ld_total < -3.029073004, penalty score =
(ld_total+3.631132995)/0.660205999
-3.029073004 <= ld_total < -2.970926996, penalty score =
(ld_total+3.029073004)/0.058146008+0.176145045+0.911927477
-2.970926996 <= ld_total < -2.368867005, penalty score =
(ld_total+3.029073004)/0.660205999+1
-2.368867005 <= ld_total < -2.029073004, penalty score = 2
-2.029073004 <= ld_total < -1.368867005, penalty score =
(ld_total+2.029073004)/0.660205999+2
-1.368867005 <= ld_total < -1.029073004, penalty score = 3
-1.029073004 <= ld_total < -0.368867005, penalty score =
(ld_total+1.029073004)/0.660205999+3
ld_total >= -0.368867005, penalty score = 4;

!Disjunctive programming algorithm for I_ah;
(-6.698970004+3.631132995)*(1-I_ah1) < ld_total+3.631132995;
ld_total+3.631132995 <= (2.301029996+3.631132995)*I_ah1;

(-6.698970004+3.029073004)*(1-I_ah2) < ld_total+3.029073004;
ld_total+3.029073004 <= (2.301029996+3.029073004)*I_ah2;

(-6.698970004+2.970926996)*(1-I_ah3) < ld_total+2.970926996;
ld_total+2.970926996 <= (2.301029996+2.970926996)*I_ah3;

(-6.698970004+2.368867005)*(1-I_ah4) < ld_total+2.368867005;
ld_total+2.368867005 <= (2.301029996+2.368867005)*I_ah4;

(-6.698970004+2.029073004)*(1-I_ah5) < ld_total+2.029073004;
ld_total+2.029073004 <= (2.301029996+2.029073004)*I_ah5;

(-6.698970004+1.368867005)*(1-I_ah6) < ld_total+1.368867005;
ld_total+1.368867005 <= (2.301029996+1.368867005)*I_ah6;

(-6.698970004+1.029073004)*(1-I_ah7) < ld_total+1.029073004;
ld_total+1.029073004 <= (2.301029996+1.029073004)*I_ah7;

(-6.698970004+0.368867005)*(1-I_ah8) < ld_total+0.368867005;
ld_total+0.368867005 <= (2.301029996+0.368867005)*I_ah8;

```

```

I_ah = (I_ah1-I_ah2)*(ld_total+3.631132995)/0.660205999+(I_ah2-
I_ah3)*((ld_total+3.029073004)/0.058146008*0.176145045+0.911927
477)+(I_ah3-I_ah4)*((ld_total+3.029073004)/0.660205999+1)
+(I_ah4-I_ah5)*2+(I_ah5-
I_ah6)*((ld_total+2.029073004)/0.660205999+2)+(I_ah6-
I_ah7)*3+(I_ah7-
I_ah8)*((ld_total+1.029073004)/0.660205999+3)+I_ah8*4;
@BIN(I_ah1); @BIN(I_ah2); @BIN(I_ah3); @BIN(I_ah4);
@BIN(I_ah5); @BIN(I_ah6); @BIN(I_ah7); @BIN(I_ah8);

!Formulation for OWA operators;
bin11*I_fl + bin12*I_ms + bin13*I_v + bin14*I_vis + bin15*I_ex
+ bin16*I_el + bin17*I_ah = r1;
bin21*I_fl + bin22*I_ms + bin23*I_v + bin24*I_vis + bin25*I_ex
+ bin26*I_el + bin27*I_ah = r2;
bin31*I_fl + bin32*I_ms + bin33*I_v + bin34*I_vis + bin35*I_ex
+ bin36*I_el + bin37*I_ah = r3;
bin41*I_fl + bin42*I_ms + bin43*I_v + bin44*I_vis + bin45*I_ex
+ bin46*I_el + bin47*I_ah = r4;
bin51*I_fl + bin52*I_ms + bin53*I_v + bin54*I_vis + bin55*I_ex
+ bin56*I_el + bin57*I_ah = r5;
bin61*I_fl + bin62*I_ms + bin63*I_v + bin64*I_vis + bin65*I_ex
+ bin66*I_el + bin67*I_ah = r6;
bin71*I_fl + bin72*I_ms + bin73*I_v + bin74*I_vis + bin75*I_ex
+ bin76*I_el + bin77*I_ah = r7;

bin11 + bin12 + bin13 + bin14 + bin15 + bin16 + bin17 = 1;
bin21 + bin22 + bin23 + bin24 + bin25 + bin26 + bin27 = 1;
bin31 + bin32 + bin33 + bin34 + bin35 + bin36 + bin37 = 1;
bin41 + bin42 + bin43 + bin44 + bin45 + bin46 + bin47 = 1;
bin51 + bin52 + bin53 + bin54 + bin55 + bin56 + bin57 = 1;
bin61 + bin62 + bin63 + bin64 + bin65 + bin66 + bin67 = 1;
bin71 + bin72 + bin73 + bin74 + bin75 + bin76 + bin77 = 1;

bin11 + bin21 + bin31 + bin41 + bin51 + bin61 + bin71 = 1;
bin12 + bin22 + bin32 + bin42 + bin52 + bin62 + bin72 = 1;
bin13 + bin23 + bin33 + bin43 + bin53 + bin63 + bin73 = 1;
bin14 + bin24 + bin34 + bin44 + bin54 + bin64 + bin74 = 1;
bin15 + bin25 + bin35 + bin45 + bin55 + bin65 + bin75 = 1;
bin16 + bin26 + bin36 + bin46 + bin56 + bin66 + bin76 = 1;
bin17 + bin27 + bin37 + bin47 + bin57 + bin67 + bin77 = 1;

@BIN(bin11); @BIN(bin12); @BIN(bin13); @BIN(bin14);
@BIN(bin15); @BIN(bin16); @BIN(bin17);
@BIN(bin21); @BIN(bin22); @BIN(bin23); @BIN(bin24);
@BIN(bin25); @BIN(bin26); @BIN(bin27);
@BIN(bin31); @BIN(bin32); @BIN(bin33); @BIN(bin34);
@BIN(bin35); @BIN(bin36); @BIN(bin37);
@BIN(bin41); @BIN(bin42); @BIN(bin43); @BIN(bin44);
@BIN(bin45); @BIN(bin46); @BIN(bin47);
@BIN(bin51); @BIN(bin52); @BIN(bin53); @BIN(bin54);
@BIN(bin55); @BIN(bin56); @BIN(bin57);
@BIN(bin61); @BIN(bin62); @BIN(bin63); @BIN(bin64);
@BIN(bin65); @BIN(bin66); @BIN(bin67);
@BIN(bin71); @BIN(bin72); @BIN(bin73); @BIN(bin74);
@BIN(bin75); @BIN(bin76); @BIN(bin77);

r1 >= r2; r2 >= r3; r3 >= r4; r4 >= r5; r5 >= r6; r6 >= r7;

I_SHI = I_fl+I_ms+I_v+I_vis+I_ex+I_el+I_ah;

```

```
!Summation of sub-index values to calculate total weighted
index score;
I_SHIr =
0.3543*r1+0.2399*r2+0.1587*r3+0.1036*r4+0.0676*r5+0.0448*r6+0.0
312*r7;

END
```